Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments

Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments

Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments

A dissertation submitted to the
University of Cambridge for the
degree of Doctor of Philosophy

James Robert Hopgood
Queens' College. November, 2000

SIGNAL PROCESSING LABORATORY
Department of Engineering
University of Cambridge

Submitted version, November, 2000.
Reprinted with minor corrections, October, 2001.

Typeset by the author with the LaTeX 2ε Documentation System, with $\mathcal{AMS}$-LaTeX Extensions, in 12/18 pt Sabon and Euler fonts.

Signal Processing Laboratory,
Department of Engineering,
University of Cambridge,
Trumpington Street,
Cambridge, CB2 1PZ, U.K.

To my parents

The Matrix is everywhere. It is all around us. Even now, in this very room. You can see it when you look out your window or when you turn on your television. You can feel it when you go to work, when you go to church, when you pay your taxes. It is the world that has been pulled over your eyes, to blind you from the truth.

'Morpheus'
The Matrix
WRITTEN AND DIRECTED BY
THE WACHOWSKI BROTHERS
WARNER STUDIOS, 1998

Declaration

I hereby certify that, except as indicated in the text, the work described in this dissertation is entirely original. It is not the result of work done in collaboration, and has not been submitted to any other University. This length of this dissertation, including footnotes, appendices, and bibliography, does not exceed 65,000 words. This dissertation contains 70 figures.

JAMES HOPGOOD
NOVEMBER 2000

Acknowledgements

I wish to express my gratitude to my supervisor, Professor Peter Rayner, for his generous advice, encouragement and support throughout this period of research. I would like to thank my colleagues in the Signal Processing Laboratory for providing a fun and intellectually stimulating environment; special thanks go to Dr. Bill Fitzgerald who has provided much advice and encouragement. I am extremely grateful to those who have proof-read and advised on sections of research including Nick Haan, Simon Hill, Peter de Rivaz and Patrick Wolfe. I would also like to thank Dave Betts at CEDAR, Cambridge, for recording a number of acoustic impulse responses which were instrumental in the latter stages of my work.

I am also grateful to the George and Lillian Schiff Foundation for providing generous financial support throughout the Ph.D., and to the Department of Engineering, Queens' College, and the Royal Academy of Engineering for providing travel funds to enable me to attend ICASSP '99 and WASPAA '99.

Finally, I would like to thank my parents for the encouragement they have given me throughout my research, and to friends who make life at Cambridge so enjoyable. Special appreciation goes to those who gave me real grief near the final stages of this thesis and insisted that I 'really should stop writing and submit the *thing*.'

Keywords

The following keywords may be useful for indexing purposes:

nonstationary stochastic processes, linear time-varying systems, nonstationary Wiener-Hopf filter, single channel/co-channel signal/speech separation and blind deconvolution, speech dereverberation, speech restoration, room acoustics, room transfer function, acoustic impulse response, ideal filter, integral transform, bifrequency transfer function, Bayesian, autoregressive processes.

Summary

In this age of instantaneous world-wide telecommunication, the demands on speech quality and intelligibility from devices such as the mobile telephone, teleconferencing systems and even the hearing aid are dramatically increasing. It is often expected that these devices will produce 'Compact Disc' quality sound and certainly that the presence of any background noise will be suppressed. However, the degradation in the quality and intelligibility of the received speech is constrained by the acoustical properties of the environment in which the audio signal is acquired. In any confined environment, the sound heard by a person is not simply the original emitted sound, but the combination of a series of an enormous number of echoes from different surfaces: the process of *reverberation*.

A considerable number of linear signal processing problems reduce to the fundamental tasks of *signal separation* and *deconvolution*. A large proportion of these problems are *blind* in the sense that neither of the source signals are known, and this substantially increases the difficulty of the problem. It is, therefore, of considerable interest to solve the problems of *single channel signal separation* and *single channel blind deconvolution* (also known as *blind dereverberation*).

This dissertation is arranged in two parts. First, the problem of determining separability criteria and achieving signal separation is considered. Separability of signal mixtures, given only one mixture observation, is defined as the identification of the accuracy to which signals can be separated. The work introduces a signal separation technique by *concatenating* the domains on which two signal classes can be represented with a finite number of basis functions. The separation of uniformly modulated signals is considered, and an example of separating chirp signals embedded in multiplicative noise is given. The second part of this dissertation considers single channel blind deconvolution, in which a degraded observed signal is modelled as the convolution of a nonstationary source signal with a stationary distortion operator. Recovery of the source signal from the observed signal is achieved by modelling the source signal as a time-varying AR process, the distortion operator by a IIR filter, and then using a Bayesian framework to estimate the parameters of the distorting filter, which can be used to deconvolve the observed signal. A further generalisation of this model using subband techniques is introduced.

Throughout the dissertation, the *philosophy* of how the nonstationary properties of the source signal allow the identification of the *interference* or *distortion* operator to be uniquely determined is discussed.

Outline Contents

Contents

List of Figures

List of Theorems, Examples and Definitions

Theorems

Examples

Definitions

List of Acronyms

AACF	accumulated autocorrelation function
ACF	autocorrelation function
AEC	acoustic echo cancellation
AIR	acoustic impulse response
AM	amplitude modulated (or amplitude modulation)
AR	autoregressive
ARCH	autoregressive conditional heteroscedastic
ARMA	autoregressive moving average
BSAR	block stationary AR
BSARMA	block stationary ARMA
BSMA	block stationary MA
CAPZ	common-acoustical pole and zero
CH	conditional heteroscedastic
DFT	discrete Fourier transform
EEG	electroencephalogram
EM	expectation-maximization
ES	evolutionary spectrum
FDM	frequency division multiplexing
FIR	finite impulse response
FM	frequency modulated (or frequency modulation)
GBTF	generalised bifrequency transfer function
GCPS	generalised cross power spectrum
GLR	generalised likelihood ratio
GMRF	Gaussian Markov random field
GPS	generalised power spectrum
HMS	harmonic magnitude suppression
HRTF	head-related transfer function
IDFT	inverse DFT
iff	if, and only if,
IGBTF	inverse generalised bifrequency transfer function

i. i. d.	independent and identically distributed
IIR	infinite impulse response
IPS	instantaneous power spectrum
KL	Karhunen-Loève (transform)
LHS	left hand side
LSE	least-squares estimate
LTI	linear time-invariant
LTV	linear time-varying
MA	moving average
MCMC	Markov chain Monte Carlo
ML	maximum-likelihood
MLE	maximum-likelihood estimate
MMAP	maximum marginal *a posteriori*
MMSE	minimum mean-square estimate
MS	mean square
MSE	mean squared error
NCAR	noncausal AR
NSHP	nonsymmetric half plane
PCM	pulse code modulation
pdf	probability density function
PSD	power spectral density
PSF	point spread function
RHS	right hand side
rj-MCMC	reversible-jump Markov chain Monte Carlo
RT	reverberation time
RTF	room transfer function
SAR	simultaneous autoregressive
SNR	signal-to-noise ratio
SRR	signal-to-reverberant component ratio
STFT	short-time Fourier transform
TDM	time division multiplexing
TVAR	time-varying AR
TVARMA	time-varying ARMA
TVMA	time-varying MA
WHF	Wiener-Hopf filter
w. r. t.	with respect to
WGN	white Gaussian noise
WSS	wide sense stationary

Nomenclature

Symbols

$\{\emptyset\}$	empty or null set.
$\mathbb{C}$	set of complex numbers.
$\mathbb{K}$	vector space.
$\mathbb{R}$	set of real numbers.
$\mathbb{R}^+$	set of positive real numbers (including zero).
$\mathbb{R}^n$	$n-$dimensional Euclidean space.
$\mathbb{Z}$	set of integers, $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$.
$\mathbb{Z}^+$	set of positive integers, $\mathbb{Z} = \{0, 1, \ldots\}$.
S or $\mathcal{S}$	matrix.
$[S]_{ij}$	element (i, j) of matrix S, *i.e.* $[S]_{ij} = S(i, j)$.
$\boldsymbol{\phi}_{-\boldsymbol{\tau}}$	all the elements of $\boldsymbol{\phi}$ except $\boldsymbol{\tau}$.
I_T	$T \times T$ identity matrix.
$\mathbf{0}$	vector of zeros.
$\mathbf{1}$	vector of ones.
diag $[\mathbf{v}]$	diagonal matrix with elements of vector $\mathbf{v}$ on leading diagonal.
$AR(P)$	AR model of order P.
$BSAR(P)$	BSAR model of order P.
$ARMA(P, Q,)$	ARMA model of order (P, Q).
$MA(Q)$	MA model of order Q.
$\triangleq$	defined as.
$\in$	'is an element of'.
$\exists$	'there exists an'.
$\Re$	real part of a complex number.
:	such that.

$\rightleftharpoons$	transform.
$\Pr(\mathcal{A})$	probability of event $\mathcal{A}$.
$\Pr(\mathcal{A} \mid \mathcal{B})$	probability of event $\mathcal{A}$ given event $\mathcal{B}$.
$p(\mathbf{x})$	pdf of $\mathbf{x}$.
$p(\mathbf{x} \mid \mathbf{y})$	conditional pdf of $\mathbf{x}$ given $\mathbf{y}$.
$\mathbf{x} \sim p(\mathbf{x})$	$\mathbf{x}$ is distributed according to $p(\mathbf{x})$.
$\mathbf{d}(t)$	bold type symbols with an associated time index: scalar stochastic processes.
$\mathbf{d}$	bold type symbols without a time index: column vectors, *i.e.* $\mathbf{d} \triangleq (d_1, \ldots, d_T)$.
$x(t)$	normal type symbols: scalar deterministic processes in Part II; scalar stochastic processes in Part III.
$\{\mathbf{x}(t)\}$	set of realisations of the stochastic process $\mathbf{x}(t)$.

Operators and Functions

$\Gamma(\cdot)$	Gamma function.
$\mathcal{IG}(\theta \mid \alpha, \beta)$	inverse Gamma distribution in θ with parameters α, β.
$\mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma)$	multivariate Gaussian distribution in $\boldsymbol{\theta}$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ.
A^{-1}	inverse of a square matrix.
$\mathbf{v}^T, A^T$	vector and matrix transpose: $[A]_{ij}^T = [A]_{ji}$.
$\operatorname{rank}[S]$	rank of matrix S.
$\langle \mathbf{u}, \mathbf{v} \rangle$	inner product of vectors $\mathbf{u}$ and $\mathbf{v}$.
$\lfloor x \rfloor$	biggest integer less than or equal to x.
$\lceil x \rceil$	smallest integer greater than or equal to x.
ln	natural logarithm function, *i.e.* to the base e.
abs	absolute value.
$\mathscr{L}$	linear operator or vector spaces.
$\operatorname{perm}\mathcal{P}$	permutation of elements in $\mathcal{P}$.
$\mathbb{I}_{\mathcal{S}}(\cdot)$	indicator function for the set $\mathcal{S}$.
$E_{p(\mathbf{x})}[f(\mathbf{x})]$	expectation of $f(\mathbf{x})$ w. r. t. the pdf of $\mathbf{x}$ (the pdf, $p(\mathbf{x})$, is omitted where it is implicit in the usage *i.e.* use $E[f(\mathbf{x})]$).
$R_{xx}(t, \tau)$	ACF of stochastic process at times (t, τ).
$\star$	convolution.
$\mathscr{F}$	Fourier transform, *i.e.* $\mathscr{F}(x(t))$ is the Fourier transform of $\{x(t)\}$.
$\mathscr{Z}$	$\mathscr{Z}-$ transform, *i.e.* $\mathscr{Z}(x(n))$ is the $\mathscr{Z}-$ transform of $\{x(n)\}$.

List of Algorithms

I

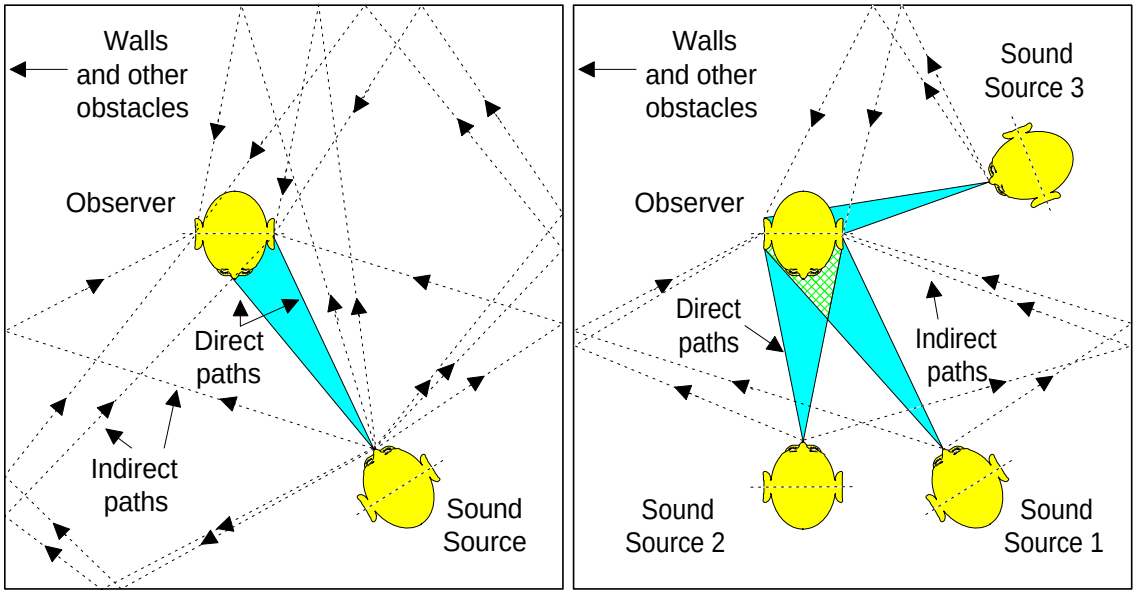
Introduction

1

Why Nonstationary Signal Processing?

IN this age of instantaneous world-wide telecommunication, the demands on speech quality and intelligibility from devices such as the mobile telephone, tele-conferencing systems and the hearing aid are dramatically increasing. It is often expected that these devices will produce near 'Compact Disc' quality sound[1] and certainly that the presence of any disturbances or distortions of speech such as background noise, echoes, clipping and other signal degradations will be suppressed or reduced to adequately low levels [116]. However, the degradation in the quality and intelligibility of the received speech is not constrained by the performance of the communication system itself but, rather, by the acoustical properties of the environment in which the audio signal is acquired. In any confined environment, the sound heard by a person is not simply the original emitted sound, but the combination of a series of a large number of echoes from different surfaces: the process of *reverberation* [191, 201], [see Figure 1.1(a)]. A person with normal hearing can, remarkably, concentrate on the original sound despite the presence of a considerable number of background disturbances. This ability is, in part, due to *binaural hearing*, which is usually referred to as the *binaural cocktail party effect*.

[1] Specifically, the quality of the transmitted speech picked up by the audio interface must, at least, be high enough to provide users with comfortable communication [116].



(a) The sound heard by an observer is a combination of the original sound, and a series of a large number of echoes from different surfaces.

(b) There will often be interfering noises, such as music or other people speaking, which complicates the reverberation problem.

Figure 1.1: *Reverberation from single and multiple sound sources.*

This effectively allows a person to concentrate on a single sound source despite the presence of many unwanted sounds such as echoes, background music and noise. In contrast, someone with a severe hearing loss perceives a degradation in sound quality and, in particular, users of hearing aids complain of being unable to distinguish one voice from another in a crowded room [99, 100, 289, 317]. This *cocktail party problem*[2] was first identified by Cherry [71] and has triggered research in widely different areas that are still relevant almost fifty years later [54]. Whilst some forms of hearing loss, for example, *conductive loss*, can be compensated for by either using a simple amplifying hearing aid or through surgery, *sensori-neural hearing loss* leads to, *inter alia*, a loss of tonal sensitivity and spatial perception [56]. A person with this impairment is unable to separate two similar frequency components in a

[2]There are several interpretations of the *cocktail party effect* in the literature. One interpretation is the so-called *cocktail party perception effect* first identified by Cherry in 1953 [71]. Another is the *cocktail party regenerative effect*, first discussed in detail by Maclean in 1959 [234], and is the all too familiar phenomenon observed when a number of people try to communicate at a cocktail party; communication becomes virtually impossible as each talker raises his individual acoustic output in order to be heard above background conversations [191, 234, 235, 289]. It is the *cocktail party perception* effect that is discussed throughout this dissertation and, for brevity, is simply referred to as the *cocktail party effect*.

complex tone, and is particularly susceptable to the effects of background noise, causing a significant decrease in intelligibility. A normal linear amplifying hearing aid cannot compensate for these losses,[3] and is unable to support binaural hearing [64, 65, 154].[4]

In many situations, such as 'hands-free' conference telephones, the sound is received by a single microphone and transferred to a remote listener as monophonic sound. As such, the listener hears a *monaural* reverberant version of the original; the advantage of the binaural cocktail party effect is lost, and the perceived quality of the speech is considerably reduced. The situation is further complicated by unwanted *background noises* which interfere with the desired sound source. These background sounds can range from environmental noise, or music, to other people speaking in the room.[5] The entire scenario is depicted in Figure 1.1(b). Another important application suffering from this effect is automatic speech recognition, where it is found more difficult to recognise reverberant speech.

These problems can be solved if the signal received at the microphone is modified before it is transmitted, so that the desired speech received by the remote user is of acceptable quality and intelligibility [252]. However, the task of modifying this signal is complicated by the fact that neither the original or background sounds, nor the acoustical properties of the room, are known beforehand. It may be possible to measure the latter in a particular laboratory, but a robust signal modification process, applicable in any given environment, cannot rely on the acoustical properties being available *a priori*. This problem is analogous to finding which two unknown numbers were added together to yield a given known number; clearly there are multiple solutions, and it is impossible to tell which is the desired solution – the problem is *degenerate*. A solution to these problems can only be found by incorporating *prior* knowledge of the characteristics of both the speech and room acoustics. The process of removing unwanted background noise is called *blind signal separation*, and the process of removing the effects of reverberation is known as *blind dereverberation* or, in the literature, *blind deconvolution* [152] or *blind system identification* [3]. After signal modification, the user should perceive only the original sound without any echoes or interfering noises, as shown in Figure 1.2.

[3] Kates [180] provides a stimulating discussion of signal processing in hearing aids, and a digital implementation of a traditional hearing aid can be found in, for example, [345].

[4] Traditional hearing aids cannot compensate for the reduction in frequency sensitivity, dynamic range, temporal resolution and spatial perception [56, 64, 65]; see Appendix A for further details.

[5] As an example, background noises in cars are in the form of *periodic noise*, due mainly to the engine, and *random noise*, such as wind, road, tires and ventilation fans [200]. Kates [178] discusses and classifies background noises that are important to consider in hearing-aid applications.

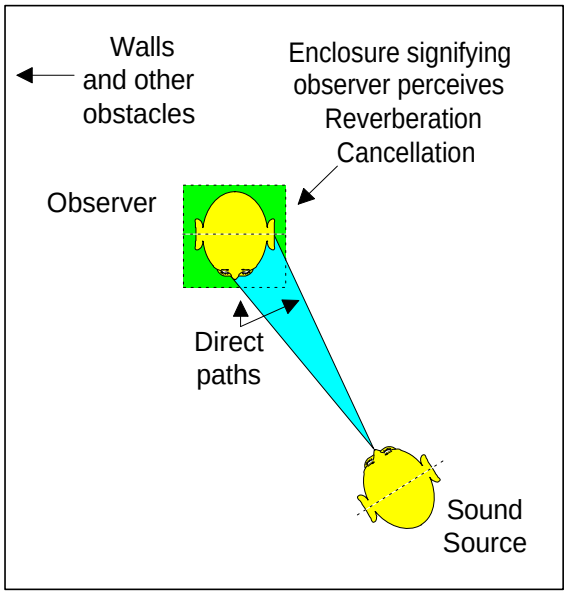


Figure 1.2: *After* reverberation cancellation *and* noise removal, *the observer perceives only the original sound without any echoes or interfering noises.*

1.1 A Simple Analogy

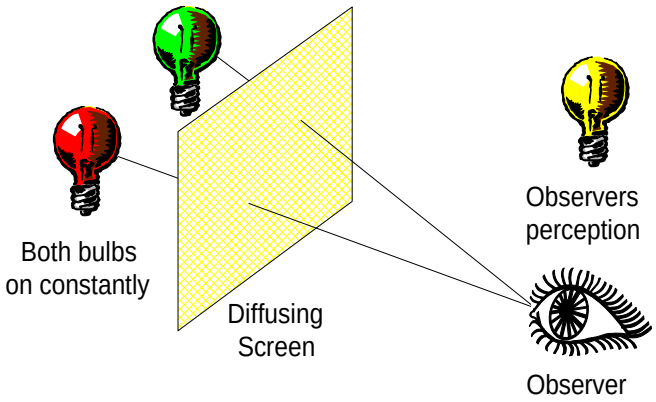
The techniques used to modify the reverberant signal measured at the microphone are studied in the branch of Engineering called *Statistical Signal Processing*. Many of the traditional methods developed in this field rely on the principle that the *statistical* properties of a system do *not* change with time, a principle called *stationarity* [286, 335]. They rely on this principle since it often reduces the complexity of a problem, and allows for the implementation of fast numerical algorithms. However, as is well known, the statistical properties of many processes, including speech, *do* change with time and, as such, many existing techniques produce poor results when applied to the blind dereverberation problem. Vaseghi and Rayner [377] discuss the effects of nonstationary signal characteristics on the performance of audio restoration systems which assume stationarity. The principle of the statistical properties of a signal changing with time, called *nonstationarity*, can be used to produce superior results for this problem, although the computational load is increased. To illustrate why this may be the case, consider the following *simple* analogy.

Suppose a red light is placed adjacent to a green light behind a diffusing screen, such that a distant observer continuously views yellow light. The observer, given no additional information other than the colour of the light he views, is asked to decide whether there is a single yellow light bulb behind the screen, *or* whether there is a red bulb *and* a green bulb.[6] If both bulbs remain on then, clearly, the observer cannot decide one way or the other; this is the shortcoming of the fact that the system is *stationary* (see Figure 1.3(a)). Now suppose the red light bulb is turned off and on, whilst the green bulb remains on continuously. It is now clear to the observer that there are two light bulbs since the light varies between yellow and green (see Figure 1.3(b)). In this case, part of the system is *stationary* and part *nonstationary*, yielding additional information that aids the observer to decide upon the number of light bulbs. Next, suppose that both the red and the green lights are turned on and off simultaneously. The observer perceives a flashing yellow bulb and, again, information regarding the system has been lost, despite the fact that the system is *nonstationary* (see Figure 1.3(c)). Finally, suppose that the red and green lights are turned on and off, but for different lengths of time. In this case the observer will see, at different times, red, green, and yellow light. In this case, both parts of the system are *nonstationary* and the observer can identify the number of light bulbs behind the screen. The additional information in this case is due to the fact that the two parts of the system have different *degrees* or *rates* of nonstationarity (see Figure 1.3(d)). By using the analogy of the mixed colours, where the light bulbs represent the sounds within a room, and the screen represents the room acoustics, it is possible to appreciate one of the basic concepts behind solving the blind dereverberation and signal separation problems.[7]

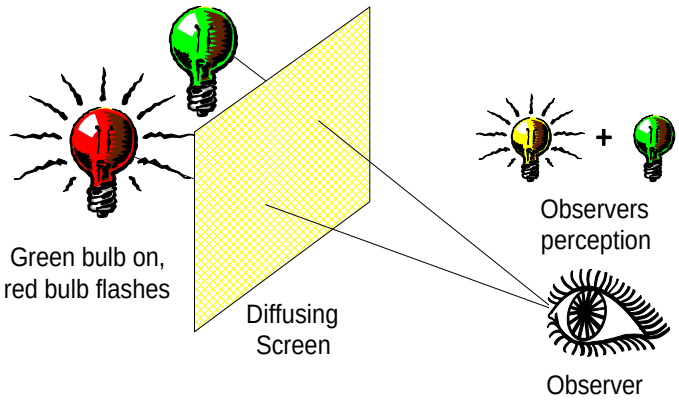
Naturally, this illustration does not give the full picture of the situation since, for example, the observer may be asked to decide whether there is a single yellow bulb behind the screen, or a red and green bulb, or a red, green *and* yellow bulb. In this case, it would appear impossible for the observer to make a decision given any of the scenarios discussed above. Therefore, knowledge that a system is nonstationary is not sufficient, and additional information is required; for example, a difference in brightness of the yellow bulb to the red and green bulbs would identify whether there is a yellow bulb or not. Alternatively, and more accurately, the observer may be told the *probability* of there being a single yellow bulb, the probability of

[6] An ambiguity arises since a mixture of red and green is perceived by the human eye as yellow.

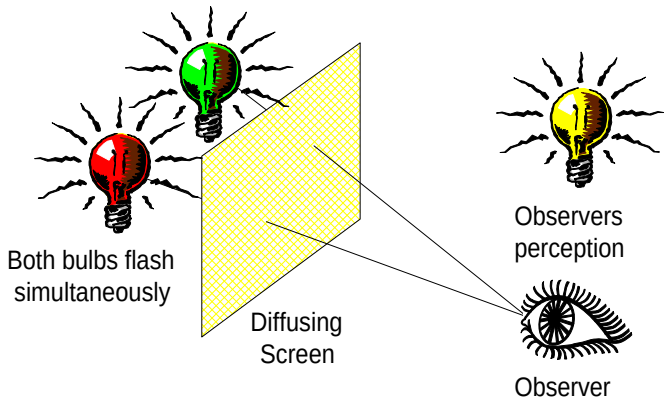
[7] The basic idea of considering images at multiple points in time is used by Law and Nguyen [209] for deblurring images although, in this case, the blurring function, or *point spread function* (PSF), is assumed to be time-varying rather than the source image, as is the case in light-bulb example.



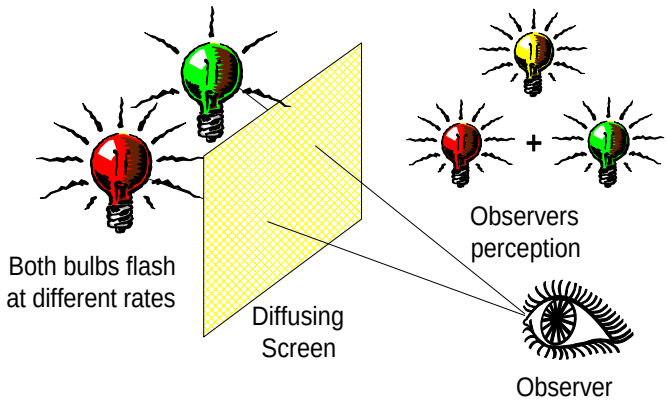
(a) System is stationary.



(b) System is part nonstationary and part stationary.



(c) System is nonstationary; each component has the same rate of nonstationarity.



(d) System is nonstationary; each component has a different rate of nonstationarity

Figure 1.3: *A simple analogy to demonstrate why nonstationarity can be used to produce superior results.*

red and green bulbs, and the probability of three bulbs. If all three colours are observed, only the probability of the number of bulbs can be stated, and the case with highest probability can be taken as the observer's conclusion of the number of bulbs. Now, the observer has been given three models for the light source behind the screen: the first model states there is a single yellow bulb, the second that there are two bulbs, and the third model states there are three. As discussed above, each model is assigned a probability of correctly representing the number of bulbs behind the screen, and this implicitly raises several questions. First, how are the models chosen? Second, how is the probability for each model assigned? Third, how many models should the observer consider and, finally, what effect does it have on the observer's conclusion if the choice of models does *not* include the correct model?

These questions must be answered systematically, avoiding *ad hoc* solutions. This research attempts to utilise *nonstationarity* in order to help solve the blind dereverberation and separation problems. There are a number of similar applications which reduce to blind deconvolution or signal separation such as speech separation, modulation schemes and multipath channel equalisation in communication systems, image processing, seismic [319] and biological data analysis,[8] where the reverberant signal is not speech, but another signal with substantially different properties [3]. In each of these applications, a suitable model must be developed, and some knowledge of the *degree* of nonstationarity must be obtained.

1.2 The 'Hands-Free' Telephone Problem

There are three major aspects, from a signal processing perspective, to the problem of degradation in speech quality in the *'hands-free' telephone problem* mentioned in the previous section: first, acoustic feedback between the loudspeaker and microphone, second, the effect of acoustic reverberation between the individual speaking and the microphone and, finally, the effect of interfering noises. A 'satisfactory' solution to this problem is a system using only conventional loudspeakers and microphones which allows a 'hand-free' telephone conversation to be held without any loss in speech quality, and allowing both speakers to move around freely in ordinary rooms. The first feature of the problem, in which the signal emitted from the loudspeaker is fed back to the microphone due to the acoustics of the room,

[8] For example, the separation of odours [159] has potential application in multimedia.

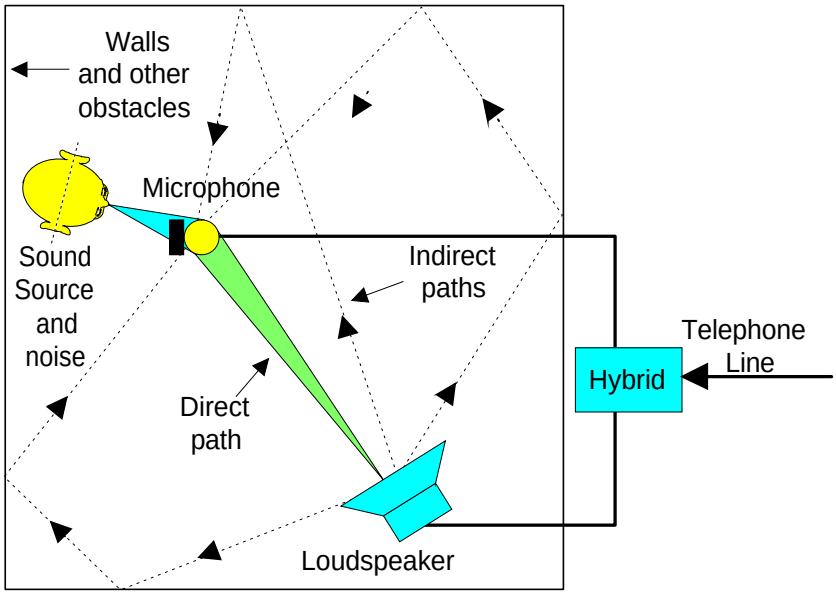


Figure 1.4: *Acoustic feedback in the 'hands-free' telephone problem.*

can lead to howling (see Figure 1.4). Since the signal driving the loudspeaker is known, howling can be eliminated by equalising the 'acoustical path' between the loudspeaker and microphone. Equalisation of this feedback path, discussed in section §6.3, is often referred to as acoustic echo cancellation (AEC), and can essentially be considered as a deconvolution problem. Although it has received much attention in the literature, with several surveys of current research in, for example, [116,147,222], a definitive solution to the AEC problem remains elusive [148]. Recent investigations into the 'hands-free' telephone problem consider AEC and *noise reduction* as a joint optimisation problem rather than distinct ones [25,136]. However, the problem of the degradation in speech quality due to interfering noises and reverberation has received less attention in the literature, and this thesis investigates solutions to these two aspects of the problem where the source sounds remain unknown.

1.3 Hearing Aids

To improve speech recognition for hearing-impaired listeners, enhanced hearing aids must preprocess the signal which its microphone receives before emitting sound

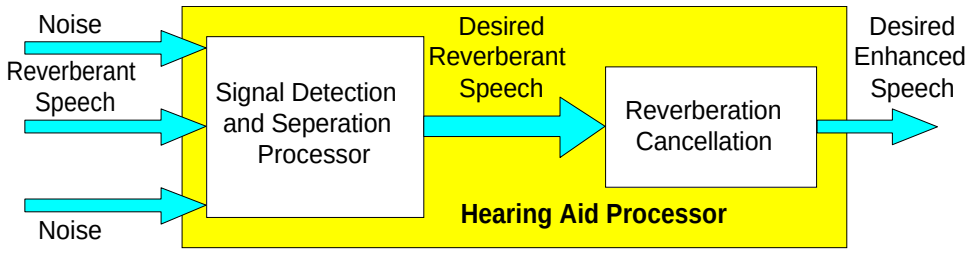


Figure 1.5: *Enhanced digital hearing aid.*

into the ear-canal. The aid should achieve this by suppressing interference coming from sources other than in the desired signal direction, and any other additional noise sources [64, 65, 178]. Microphone arrays, which can detect the direction of sound sources, can, in part, compensate for the inability of the hearing-impaired to resolve acoustical cues [56, 157, 177, 179, 224, 317]. However, the effect of reverberation, which lacks a specific direction of arrival, is a significant cause of degradation in speech intelligibility and remains an unsolved problem. Figure 1.5 shows the general form of an enhanced digital hearing aid.

Acoustic feedback oscillation, which is a problem for *'hands-free' telephone applications*, is also a problem in hearing aids and, to prevent howling, often prevents the attainment of an adequate amount of gain for some users [176, 178, 181, 182, 247, 396]. The acoustic feedback path includes the effects of the hearing-aid amplifier, receiver and microphone, as well as the acoustics of the *vent* or *leak* of the hearing aid [180, 181]. An adaptive filter is often used to cancel the acoustic and mechanical feedback picked up by the microphone, thus allowing more gain in the hearing aid. Kaelin *et al.* [171, 401] discuss a typical implementation of a digital hearing aid with *loudness compensation* and acoustic echo cancellation.

1.4 Nonstationarity

Signal processing techniques over the past three decades have been dominated by the constraint of *stationarity* – the assumption that the statistics of a process or system do not change with time – a dominance which can be attributed to the simplification of problems arising from such an assumption. The notion of sta-

tionarity is appealing as many processes are endowed with the property of ergodicity [286, 335]. Ergodicity allows quantities defined as ensemble averages, such as, for example, autocorrelation functions, to be estimated from a single realisation of the process by calculating time domain averages. However, the estimation of the autocorrelation function of a nonstationary processes is difficult, often requiring multiple data records [22] which may not be available, a problem addressed in section §4.3. Consequently, nonstationarity is usually regarded as an undesirable feature, inasmuch as it significantly increases the complexity of a problem. Nevertheless, over the past few years it has been recognised that nonstationarity can actually be a useful feature. The purpose of this section is to discuss how nonstationarity can be utilised to produce superior results, both in existing problems attempted using the stationarity assumption, and in previously intractable problems. Section §1.4.1 begins by briefly considering the natural aspects of human hearing which effectively utilise the nonstationarity inherent in natural processes.

1.4.1 Nonstationarity in Human Hearing

The human hearing mechanism is very sophisticated and, as already mentioned, people with normal hearing can separate desired speech from unwanted interfering speech with a remarkable degree of intelligibility, using the binaural cocktail party effect. Most *natural* sounds can be identified and separated by the hearing mechanism through exploitation of *spectral*, *temporal* and *spatial* features [287]. The spectrum of a sound is important for determining its information carrying characteristics, while spatial and temporal variations provide important cues for localising sound sources.[9] When these cues are unavailable, the brain finds it difficult, or near impossible, to separate jumbled speech received over, for example, monaural air-traffic control channels, conference telephones, or hearing aids. These spectral, temporal and spatial characteristics are important pointers to the information that should be incorporated into a blind dereverberation or signal separation algorithm.

1.4.1.1 Spectral Characteristics

The rôle of spectral characteristics is directly analogous to the light bulb analogy discussed in section §1.1, and this example will be recast in terms of power spectral (Fourier) components, or even just pure tones. Suppose it is desirable to separate two signals; there are several obvious different scenarios in which this is possible:

[9] See Appendix A for a further discussion on the psychology of hearing.

- **Non-overlapping spectral components:** Suppose the power spectra of the signals do not overlap in the Fourier domain. Separation can be achieved when the frequency bands of the spectra are known *a priori*; this is akin to viewing the light bulbs from behind the diffusing screen where individual colours are observed.

- **Overlapping stationary and nonstationary spectral components:** Suppose the power spectrum of one of the signals is broad-band, and the second signal consists of a single frequency that overlaps the spectrum of the first. If both signals are stationary, it is impossible, without significant prior knowledge, to separate them. However, suppose the second signal is known to have an unknown time-varying amplitude. In an 'ideal' time-frequency spectrum of the mixture, where a fixed-frequency amplitude-varying sinusoid appears as a straight line with time-varying amplitude, the time-varying power spectral component can be attributed to the second signal.[10] This is comparable to one of the light bulbs flashing while the other remains constantly on.

1.4.1.2 Temporal Characteristics

The temporal characteristics of unknown signals play an important part in their identification. If the mixture of two signals is observed and it is known that the signals do not overlap in the time domain then, provided the instances at which the signals are 'switched on' are known, the signals can be identified. This is equivalent to the case when the light bulbs flash at different instances.

1.4.1.3 Spatial Characteristics

The human hearing mechanism uses spatial localisation of sounds in order to reduce the intrusion of noise. When a sound is localised, a listener will often turn their head to use the natural attenuation of the outer ear and skull to reduce the distracting effect of other sounds.[11] Moreover, visual cues, such as lip reading, gestures, and the like [71, 128], aid the listener to identify the desired source. This additional tuning of the auditory system aids concentration and is characteristic of the incorporation of *prior knowledge*. Although the desired sound is initially unknown, these tuning mechanisms influence the brain's focus on that sound, and

[10] There, naturally, will be an ambiguity in the absolute signal amplitude.

[11] This movement also acts to modify the *head-related transfer function* (HRTF) to reduce reverberant effects in the cranium (see page 139).

serve to eradicate the effect of noise. As such, it is clearly important to take advantage of any prior knowledge in attempting to separate or dereverberate a signal.

1.4.2 Nonstationarity in Signal Processing

Over the past few years, it has been acknowledged that the principle of nonstationarity, usually regarded as an undesirable feature rendering many problems intractable, can actually be useful in improving the performance of many signal processing solutions. As an example, a least-squares solution is the conventional method for identifying the transfer function of an unknown linear system given by:

$$y(t) = h(t) \star x(t) + n(t), \quad \forall t \in \mathscr{T} \subset \mathbb{R} \tag{1.1}$$

where $h(t)$ represents the impulse response of the *linear time-invariant* (LTI) system, $\star$ denotes the convolution operation, $\{x(t)\}$ and $\{y(t)\}$ are the observed input and output processes, and $\{n(t)\}$ represents additive observation and modelling noise. The least-squares approach is equivalent to identification of the frequency response of the system by calculating the ratio of the cross-power spectrum of the input and output to the power-spectrum of the input signal. However, if the additive noise, $n(t)$, is correlated with the input process, least-squares suffers from a severe bias effect. As such, Shalvi and Weinstein [322] present a solution which exploits the nonstationary features of the signals to circumvent these biasing effects. Nonstationary inputs are also useful for *nonminimum-phase* system identification (see Chapter 6), and are used by Al-Shoshan and Chaparro [7] and Xia [402] in which the *time-frequency domain* is exploited (see section §3.2.1). Further examples of the improvement in signal processing solutions from the exploitation of nonstationarity occur in denoising [153], and multichannel signal separation [389] (see section §2.2), where more robust separation solutions are obtained than when using the assumption of stationarity [4, 5].

1.5 Objectives of the Thesis

This dissertation is concerned with utilising nonstationary signal processing to improve the quality and intelligibility of a monaural audio signal which consists of

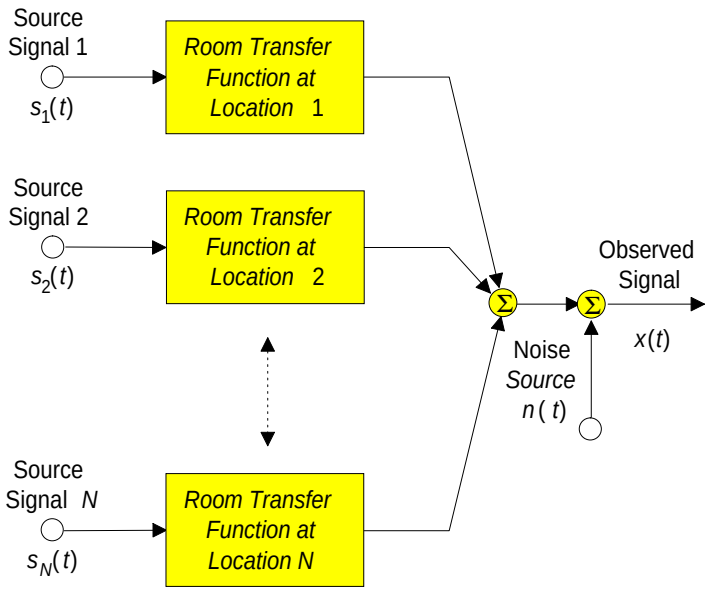


Figure 1.6: *This signal model represents the monaural audio signal received by the user for the scenario shown in Figure 1.1.*

multiple reverberant source signals. The general physical scenario for this problem, shown in Figure 1.1, can be represented by the signal model in Figure 1.6. Although the source signals can be recovered from the observed signal by solving the entire problem using, for example, a Bayesian approach, the objective of this dissertation is to investigate how the utilisation of nonstationarity signal processing can lead to improved results. Whilst in [25] it has been proposed that *noise reduction* and *acoustic echo cancellation* should be performed jointly for 'hands-free' telephones, the general *blind multi-source reverberation problem* is divided into two separate distinct problems and are considered separately since, unlike acoustic echo cancellation, all the source signals are unknown.

- First, in Part II of this thesis, single channel signal separation using linear filters is investigated. Although a Bayesian methodology can be used for source estimation from a mixture of signals, in this instance the approach does not advance a deeper understanding of the underlying philosophy of how nonstationarity facilitates the separation of seemingly inseparable signals.

- Second, in Part III, single channel blind deconvolution is considered, with application to dereverberation of speech signals in acoustic environments. In this case, the Bayesian approach is used for channel estimation since a thorough appreciation of the rôle of nonstationary effects can be obtained by an

analysis of the underlying mechanics of the Bayesian approach.

These two approaches are fundamentally different in technique, yet have similarities in their dependence on nonstationary effects.

1.6 Overview of Dissertation

A detailed outline of the remaining chapters in this dissertation is presented below. Some of the material in Chapters 3 and 4 appeared in [162], and has been submitted in the following articles: [163, 164]. Material in Chapter 8 has appeared in [161].

Chapters containing original and significant contributions are denoted by a [*] *in the following list.*

1.6.1 Part II: Single Channel Signal Separation using Linear Filters

Chapter 2 provides a general introduction to the problem of single channel signal separation, defines signal separability, and gives a brief review of existing approaches to the problem.

Chapter 3 [*] gives an overview of the Wiener-Hopf filter and introduces the fundamental concepts needed to analyse nonstationary stochastic processes and linear time-varying systems. This chapter also introduces the concept of the *generalised power spectrum*.

Chapter 4 [*] reviews the so-called *ideal filter*, and presents a spectral solution for the nonstationary Wiener-Hopf filter. These results naturally lead to a criterion for the separation of nonstationary signals with perfect precision, and is a natural extension to the condition for separating stationary signals. The chapter introduces a signal separation technique by *concatenating* the domains on which two signal classes can be represented with a finite number of basis functions. Examples are given of the separation of classes of signals which would 'classically' be considered inseparable. These include uniformly modulated signals which can model multiplicative noise and, in particular, the separation of *chirp signals* which occur in many practical applications such as radar and speech.

Chapter 5 summarises the results in Part II.

1.6.2 Part III: Single Channel Blind Deconvolution

Chapter 6 introduces some basic theoretical properties of acoustics that are important for understanding why particular models are used in this work, and the problems associated with attempting to dereverberate speech. The suitability of some well-known modelling techniques in signal processing for the representation of room acoustics, the robustness of these models to variations in the source and observer position, and the effect of parameter variation on the accuracy of the model are all discussed. Finally, some existing approaches to the enhancement of reverberant speech are reviewed.

Chapter 7 provides a more general introduction to the problem of single channel blind deconvolution. The chapter discusses linear input–output modelling to represent systems and nonstationary stochastic processes in terms of finite-order linear models. The Bayesian paradigm is introduced as a means of robust parameter estimation.

Chapter 8 [*] introduces the basic model which facilitates blind deconvolution by utilising the nonstationary properties of the system. A 'physical' interpretation of the underlying mechanism at work in the process is presented, as is an example of deconvolving a source signal with a distortion filter. Finally, some fundamental issues regarding the estimation of unknown system parameters is considered and, more importantly, in light of the additional knowledge gained from taking advantage of the nonstationarity in the system, the question regarding the accuracy to which these parameters must be estimated is raised.

Chapter 9 [*] generalises the model presented in Chapter 8 using subband techniques which, not only reduces numerical complexity and errors resulting from the modelling of long impulse response functions, but facilitates dereverberation by identifying which bands of the room transfer function are nonminimum-phase. The issues associated with such a generalisation are intimately related to the issues raised in Chapter 8, and are discussed further.

Chapter 10 summarises the results presented in Part III.

1.6.3 Part IV: Conclusions, Appendices and Bibliography

Chapter 11 concludes the dissertation by summarising the main issues which have arisen in these investigations, and outlines the direction that future research in this area could follow. Specific details are given for extending the theory of Part II to deal with time-frequency distributions, and details are given for modifying the blind deconvolution methodology to deal with two-dimensional signals, such as images.

Appendices A to E contain further details and proofs of results used throughout the thesis which have been omitted from the main body of the text for clarity.

II

Single Channel Signal Separation of Random Signals using Linear Time-Varying Filters

2

Introduction to Signal Separation

IN the first part of this dissertation, a linear filtering approach to the problem of single channel signal separation is discussed. When the auto- and cross-correlations of two source signals are known *a priori*, the problem is solved optimally using the Wiener or Kalman filter. If the source signals are nonstationary, these correlation functions can be very difficult to estimate, and the Wiener or Kalman filters are difficult, if not impossible, to design. Therefore, when only partial information of the auto- and cross-correlation functions are known, it is relevant to ask what constraints on the unknown characteristics of the signals are required to achieve a given degree of separability; this is referred to as the *signal separability* problem.

In order to answer the separability problem, it is necessary to view the structure of signals from a more useful perspective than that of the time domain and, as such, the *generalised integral transform*, also known as a *generalised signal decomposition*, is introduced. It is also of great interest to compare the general solution of the Wiener filter with the so-called 'ideal filter', which yields perfect signal separation for certain classes of signals given only partial information pertaining to the correlation functions. Methods are discussed for determining the class for which

a given signal is a member. In particular, for the special class of uniformly modulated signals, constraints are derived on the source signal characteristics which yield 'perfect separation'. Some of the more popular methods for nonstationary signal separation discussed in the literature are reviewed in section §2.2. Single channel signal separability criteria for nonstationary signals are, however, less forthcoming in the literature.

2.1 Signal Modulation and Separation

There are many applications where the mixture of two or more signals is observed and, at each time instance, there is only a *single* observation of the mixture in the time domain. Applications range from taking multiple measurements in seismology to the design of modulation schemes in telecommunications. There are various modulation schemes in communication theory that have the property that different signals may be modulated such that their Fourier spectra overlap and recovery of signals is still possible, for example, in spread spectrum modulation.[1] However, some schemes do facilitate perfect recovery: for example, consider observing the sum of the signals in a quadrature modulation scheme in the time domain,[2] which can be 'perfectly' separated using the quadrature demodulator shown in Figure 2.1. The resultant signals are disjoint in neither the temporal nor Fourier domain; however, as shown in section §4.5, there does, in fact, exist a spectral domain in which their representations are disjoint. There is, therefore, a close link between classes of signals which can be separated and examples from existing modulation schemes. However, whilst in modulation schemes the modulated carrier is always known, such a distinguishing class feature is unknown in signal separation problems. The general distinguishing features of the signal classes *must* be known, *otherwise* it would be impossible to assess when a separation algorithm is performing correctly.

In this contribution, it is assumed that the signal classes exhibit some general distinguishing features, and that these features are known *a priori*. Identification and accurate estimation of these properties and any corresponding parameters is treated as a separate problem and left for future work. Once these properties are known, the techniques discussed in this dissertation can be applied.

[1] Albeit in some cases the separation may not be 'perfect'.
[2] See section §4.6.1 for more details.

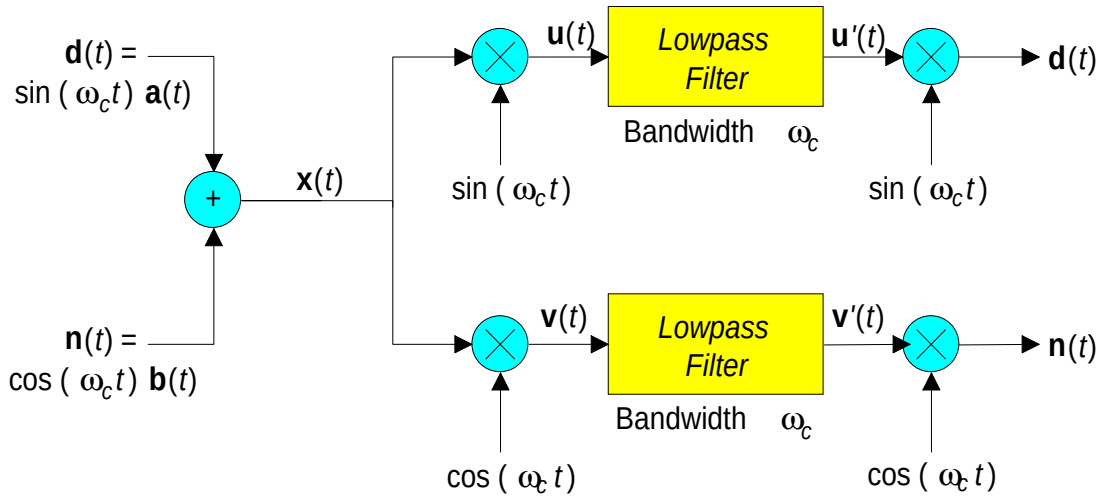


Figure 2.1: *Quadrature demodulator.*

The separation of n signals given n sensors has received much attention in recent research, for example, see Weinstein *et al.* [389], Molgedey and Schuster [254], Chan [61], Ahmed *et al.* [4, 5] and Godsill and Andrieu [119]. However, a *general* approach to signal separation given one sensor has received much less attention. In section §4.5, the separation of mixed signals in the form:[3]

$$\begin{aligned} \mathbf{d}(t) &= \int_{\mathcal{T}} h_d(t,\tau)\, \mathbf{a}(\tau)\, d\tau \\ \mathbf{n}(t) &= \int_{\mathcal{T}} h_n(t,\tau)\, \mathbf{b}(\tau)\, d\tau \end{aligned} \tag{2.1a}$$

where $t \in \mathcal{T} \subset \mathbb{R}$, is investigated. It is assumed that the *unknown* signals $\{\mathbf{a}(t),\ t \in \mathcal{T}\}$ and $\{\mathbf{b}(t),\ t \in \mathcal{T}\}$ are bandlimited to $\pm\omega_c$ and the *known* functions $h_d(t,\tau)$ and $h_n(t,\tau)$ are such that the resultant signals, $\mathbf{d}(t)$ and $\mathbf{n}(t)$, are overlapping in the Fourier domain.[4] Thus, traditional bandpass filtering is no longer adequate. The observed signal is given by:

$$\mathbf{x}(t) = \underbrace{h_d(t,\tau) \star \mathbf{a}(\tau)}_{\mathbf{d}(t)} + \underbrace{h_n(t,\tau) \star \mathbf{b}(\tau)}_{\mathbf{n}(t)} \tag{2.1b}$$

[3] In the first part of this dissertation, bold symbols, with an associated time index, represent scalar stochastic processes, *e.g.*, $\mathbf{x}(t)$; bold symbols without a time index are vectors, *e.g.*, $\mathbf{x}$. Normal type symbols represent scalar deterministic processes.

[4] In order to avoid the trivial case where a conventional bandpass filter in the Fourier frequency domain can separate the signals.

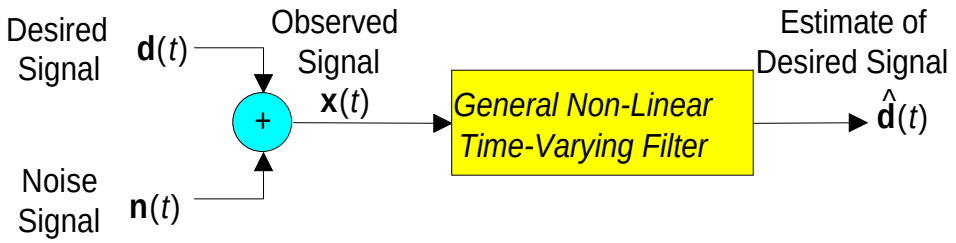


Figure 2.2: *Single channel signal separation problem.*

Modelling a stochastic process as a known time-varying filtration of some other stochastic process in the form of equation (2.1a) encapsulates quadrature modulation and spread spectrum methods, as well as simple models of speech sounds (see section §4.6.2).

2.1.1 Signal Separability

The *separability* of signal mixtures will be investigated given, at each time instance, only *one* observation of the mixture in the temporal domain. In this context, separability means identifying whether the mixture of two signals can be separated to a given degree of accuracy. The problem of actually separating the signals is referred to as *signal separation*. Formally, the problem is defined as follows:

Definition 1 (Signal Separability). Suppose a desired signal, $\mathbf{d}(\mathrm{t})$, is corrupted by an additive noise signal, $\mathbf{n}(\mathrm{t})$, such that the observation, $\mathbf{x}(\mathrm{t})$, of the desired signal is given by $\mathbf{x}(\mathrm{t}) = \mathbf{d}(\mathrm{t}) + \mathbf{n}(\mathrm{t}), \forall \mathrm{t} \in \mathscr{T} \subset \mathbb{R}$. The problem of separability is to determine conditions on $\mathbf{d}(\mathrm{t})$ and $\mathbf{n}(\mathrm{t})$ such that an estimate of the desired signal, $\hat{\mathbf{d}}(\mathrm{t}), \forall \mathrm{t} \in \mathscr{T}$, can be obtained, from $\mathbf{x}(\mathrm{t})$, to a given degree of accuracy using a general nonlinear time-varying filter. Figure 2.2 shows the general relationship between these signals.

2.1.2 Separation Techniques

A general separability criterion is very difficult to derive, since any such result would be a function of the separation method. To illustrate this, consider the separability of two signals that do not overlap in the time domain. These signals can be separated using a temporal switch, which is a special case of a linear time-varying (LTV) filter. Separability criteria for these signals depend on the times

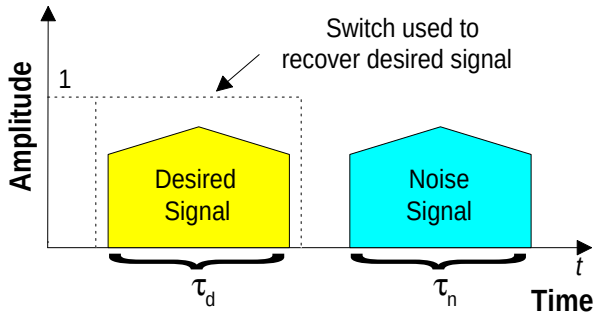


Figure 2.3: *Separability of signals in the time domain.*

for which they have non-zero components; *i.e.* $\tau_{\mathrm{d}} = \{\mathrm{t} \in \mathbb{R} : \mathbf{d}(\mathrm{t}) \neq 0\} \subset \mathbb{R}$ and $\tau_{\mathrm{n}} = \{\mathrm{t} \in \mathbb{R} : \mathbf{n}(\mathrm{t}) \neq 0\} \subset \mathbb{R}$ (see Figure 2.3). However, these signals cannot be separated using a linear time-invariant (LTI) filter and, therefore, in such a case, no separability criteria would exist and separation would appear to be impossible. It is well known that certain classes of signals can be separated using nonlinear filters (*e.g.* a pulse code modulation (PCM) decoder [418]) but, since there does not currently exist a general theory to deal with nonlinear filters, separability criteria are derived assuming that separation is achieved using LTV filters. However, even if signals are not separable under the criteria derived here, they are not necessarily inseparable: it merely implies that they are not separable using deterministic LTV filters designed under the specific choice of *cost function* discussed in section §3.1.1. For example, Arakawa *et al.* [16, 277] suggest a nonlinear digital filter for separating the nonstationary component, such as spikes and 'crackles', from electroencephalogram (EEG) data in which the stationary component is represented by an autoregressive (AR) process; however, since the 'degree of separation' depends on the nonlinearity chosen, some additional prior knowledge regarding the nonstationarity is required. A nonlinear thresholding method for separating the nonstationary component is discussed in [311], while Godsill and Rayner [121] (see also [117]) discuss an approach for removing impulsive noise from AR processes using Bayesian techniques.

2.1.3 Prior Knowledge

Signal separation with one observation sensor can only be achieved by exploiting prior knowledge of the signal structure, since the separation problem is inherently under-constrained: at each time instant, there are two unknowns, and one equa-

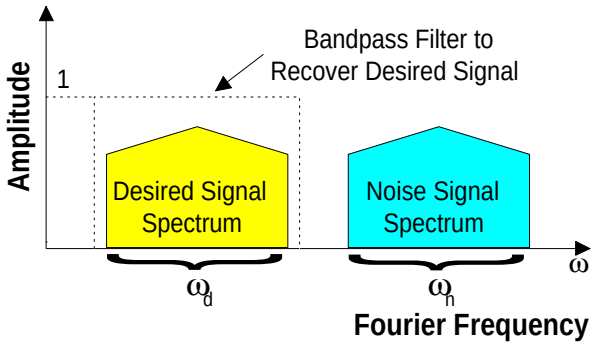


Figure 2.4: *Separability of signals in the Fourier domain.*

tion. Moreover, it is desirable that any prior knowledge necessary for signal separation can be specified such that it is common to a 'class' of stochastic processes. This prior information must somehow be estimated either from the signal, or from some known signal model. However, it is again noted that whilst the problem of determining constraints on these distinguishing features for achieving separation to an arbitrary accuracy is addressed in this dissertation, the problem of estimating them is not and is left as future research (see Chapter 11). Furthermore, only a limited amount of prior information is required for separation. So, for example, it is well known that stationary stochastic signals are 'perfectly' separable if their power spectra do not overlap in the Fourier domain. Separation can be achieved using a bandpass filter, as shown in Figure 2.4, when the frequencies in the passband, $\omega_d = \{\omega \in \mathbb{R} : |\mathbf{D}(\omega)| > 0\} \subset \mathbb{R}$ and $\omega_n = \{\omega \in \mathbb{R} : |\mathbf{N}(\omega)| > 0\} \subset \mathbb{R}$, are known *a priori*, where $\mathbf{D}(\omega)$ and $\mathbf{N}(\omega)$ are the spectral components of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ respectively. It is not, however, necessary to know the particular set of values of $\mathbf{D}(\omega)$ and $\mathbf{N}(\omega)$. Similarly, as noted in the previous section, nonstationary stochastic signals which are non-overlapping in the time domain can be perfectly separated if the set $(\tau_d, \tau_n) \subset \mathbb{R} \times \mathbb{R}$, $\tau_d \cap \tau_d = \{\emptyset\}$, is known.

Since finite bandwidth non-overlapping signals in the Fourier domain can be separated using a LTI bandpass filter, and finite duration non-overlapping signals can be separated using a LTV filter, *i.e.* a switch, this raises the question whether there exists an arbitrary signal domain such that the representation of two classes of signals are disjoint and, therefore, whether signal separation can be achieved using a generalised bandpass filter on that domain. Such a filter may be neither as simple as a switch nor conventional bandpass filter, but somewhere in between these two extremes.

2.1.4 Power Spectra, Transfer Functions, and Ideal Filters

If a generalised signal domain can be found such that $\{\mathbf{d}(t)\}$ and $\{\mathbf{n}(t)\}$ have components lying in disjoint regions, a generalisation of the *bandpass filter*, generally associated with filtering in the Fourier domain, can be developed. This generalised filter, first proposed by Zadeh [413–415, 418], and later used by Tamvaclis [346, 347], is called an *ideal filter* and, given the observed signal $\{\mathbf{x}(t)\}$, can be used to recover $\{\mathbf{d}(t)\}$. The concept of a generalised signal domain to determine the separability of stochastic signals is introduced, and the relationship between this idea and the concept of power spectra and transfer functions, tools well understood in stationary system analysis, are outlined in detail.

2.2 Review of Approaches to Single Channel Signal Separation

The necessity of *prior* knowledge regarding the characteristics of the desired and noise signals was discussed in the previous section. The requirement of additional knowledge for single channel signal separation is well known, and indeed plays an important rôle in estimation theory and optimum filtering [335, 353]. This section reviews some of the various approaches used in existing single channel separation techniques, and motivates the development of a general filtering theory for signal separation, and highlights the fundamental differences between previous approaches to the problem and that presented in this work.

2.2.1 Deterministic Signals in Noise

The problem of single channel signal separation may be considered as the *estimation* of a signal in the presence of noise. A special case of this problem leading to simple solutions is that of a deterministic signal in noise, a typical problem in radar applications, where a signal of known form is reflected from a distant target. The received signal is the sum:

$$\mathbf{x}(t) = \mathrm{d}(t) + \mathbf{n}(t)$$

where the expected value of $\mathbf{n}(t)$ is zero, *i.e.*, $E[\mathbf{n}(t)] = 0$, $d(t)$ is a shifted and scaled version of the transmitted signal $\hat{d}(t)$, and $\mathbf{n}(t)$ is a noise signal. Therefore,

$$\exists\, a \in \mathbb{R},\ \exists\, \tau \in \mathscr{T} \,:\, d(t) = a\,\hat{d}(t-\tau)$$

It is assumed that the transmitted signal, $\hat{d}(t)$, is known, and the problem is to establish the presence and location of the reflecting object, which can be determined from the amplitude, a, and time delay, τ, of the received signal, $d(t)$. Hence the problem is to estimate $d(t)$ given knowledge of $\hat{d}(t)$. Some characteristic of the noise signal is required to solve the problem for, if it remained unknown, it is possible that the noise could be of similar form to $\hat{d}(t)$, in which case it would be impossible to determine whether the estimate of $d(t)$ is accurate or not. To obtain a tractable solution, $\mathbf{n}(t)$ is assumed to be a wide sense stationary (WSS) process with known power spectrum, $N(\omega)$. Given these signal characteristics, a LTI filter with impulse response, $h(t)$, is applied to the observed signal, $\mathbf{x}(t)$, with $h(t)$ chosen so as to maximise the signal-to-noise ratio (SNR) [286]. This situation is of limited interest in the present work, as effectively one of the signals is known *a priori*.

2.2.2 Model-Based Signal Separation Techniques

It is common to classify signals using model-based signal separation techniques [117, 120, 122, 123, 333]; a discussion of commonly used parametric models can be found in section §7.5. A very common model is the autoregressive (AR) model (see section §7.5.1.2, [237]), which is extremely useful in modelling speech processes: the speech is modelled as a filter, corresponding to the vocal-tract transfer function, excited by either a quasi-periodic train of impulses, corresponding to voiced excitation, or a random noise source, corresponding to unvoiced excitation. In the following example, presented in Godsill and Tan [117, 123], the signal and noise are modelled using AR processes, and a maximum marginal *a posteriori* (MMAP) estimate is proposed as a solution for separating the sampled signals $\{\mathbf{x}_1(t)\}$ and $\{\mathbf{x}_2(t)\}$, where t takes on integer values, $t \in \mathbb{Z}$. These two signals are modelled using an AR process described by:

$$\mathbf{x}_i(t) = \sum_{p=1}^{P_i} a_i(p)\,\mathbf{x}_i(t-p) + \mathbf{e}_i(t), \quad i \in \{1,2\},\ t \in \mathbb{Z} \tag{2.2}$$

where $e_i(t) \sim \mathcal{N}\left(e_i(t) \,|\, 0,\ \sigma_i^2\right)$ is a zero mean white Gaussian noise (WGN) process with excitation variance σ_i^2. The observed signal is given by:

$$x(t) = x_1(t) + x_2(t) \tag{2.3a}$$

Now suppose the AR parameters, excitation variances and model orders for the two AR processes, given by $\boldsymbol{\theta} = \{a_i(p),\ p \in \{1, \ldots P_i\},\ \sigma_i^2,\ P_i,\ i \in \{1, 2\}\}$, are known. If these parameters are unknown, as is likely in real problems, they can be removed within the *Bayesian* framework, introduced below and discussed further in section §7.7, where the unknown *nuisance* parameters are *marginalised* out of the joint probability density function (pdf) such that a solution can still be found, albeit, numerical techniques may be required. Nevertheless, to illustrate the underlying philosophy of this technique, the parameters are assumed to be known in this example. Taking a vector of observations, $\mathbf{x}$, containing T samples, equation (2.3a) may be written as:

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$$

with $\mathbf{x}_i$ written as $$\mathbf{x}_i = A_i\, \mathbf{e}_i, \quad i \in \{1, 2\} \tag{2.3b}$$

and where A_i is the *prediction error matrix* containing AR coefficients from the i-th model. The matrix is appropriately defined so as to obtain the required relationship in (2.3b). Under the assumption that the signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, the pdf of $\mathbf{x}_i$ can be written using the probability transformation [286] between $\mathbf{e}_i$ and $\mathbf{x}_i$ as:[5]

$$p_{\mathbf{x}_i}\left(\mathbf{x}_i \,|\, \boldsymbol{\theta}\right) = \frac{1}{\left(2\pi\sigma_i^2\right)^{\frac{T-P_i}{2}}} \exp\left(-\frac{\mathbf{x}_i^T\, A_i^T\, A_i\, \mathbf{x}_i}{2\sigma_i^2}\right), \qquad i \in \{1, 2\} \tag{2.4a}$$

The probability for the observed data conditional upon $\mathbf{x}_1$ is given by

$$p\left(\mathbf{x} \,|\, \boldsymbol{\theta},\, \mathbf{x}_1\right) = p_{\mathbf{x}_2}\left(\mathbf{x} - \mathbf{x}_1 \,|\, \boldsymbol{\theta}\right) \tag{2.4b}$$

and the posterior distribution for $\mathbf{x}_1$ is obtained using Bayes' rule:[6]

$$p\left(\mathbf{x}_1 \,|\, \boldsymbol{\theta},\, \mathbf{x}\right) = \frac{p\left(\mathbf{x} \,|\, \boldsymbol{\theta},\, \mathbf{x}_1\right)\ p_{\mathbf{x}_1}\left(\mathbf{x}_1 \,|\, \boldsymbol{\theta}\right)}{p\left(\mathbf{x} \,|\, \boldsymbol{\theta}\right)} = \frac{p_{\mathbf{x}_2}\left(\mathbf{x} - \mathbf{x}_1 \,|\, \boldsymbol{\theta}\right)\ p_{\mathbf{x}_1}\left(\mathbf{x}_1 \,|\, \boldsymbol{\theta}\right)}{p\left(\mathbf{x} \,|\, \boldsymbol{\theta}\right)} \tag{2.4c}$$

[5] See section §7.7.3 for further details regarding this probability transformation.
[6] See section §7.7 for a complete description of this methodology.

Hence, maximising the numerator in (2.4c), using the expressions from (2.4a), gives the MMAP (see section §7.7.1) estimate of $\mathbf{x}_1$, $\mathbf{x}_1^{MMAP}$, given by:

$$\left(\frac{\mathrm{A}_1^{\mathsf{T}}\mathrm{A}_1}{\sigma_1^2}+\frac{\mathrm{A}_2^{\mathsf{T}}\mathrm{A}_2}{\sigma_2^2}\right)\mathbf{x}_1^{MMAP}=\frac{\mathrm{A}_2^{\mathsf{T}}\mathrm{A}_2}{\sigma_2^2}\mathbf{x} \tag{2.5}$$

If the excitation noise is coloured, but with a known distribution, a MMAP estimate can be found using a similar method, although analytic results may not be possible. This separation technique does not yield perfect separation and, for example, in the case of $\mathrm{A}_1 = \mathrm{A}_2$ and $\sigma_1 = \sigma_2$, the MMAP estimate is merely the average of $\mathbf{x}_1$ and $\mathbf{x}_2$. 'Good' separation is possible with a MMAP estimate if the pdfs are 'well separated' in the probability space.

Part III of this dissertation uses the Bayesian framework to investigate the *blind deconvolution* problem. It will be seen that an intuitive explanation is forthcoming of how the use of nonstationarity allows improved parameter estimates to be obtained. Unfortunately, such an elegant explanation is not available for the signal separation problem. Nevertheless, the techniques discussed in Part III can be applied to signal separation, as has been investigated for the multichannel case in, for example, [119]. Bayesian methods have also been used in the related problem of estimating sinusoids in noise [14, 15].

2.2.3 Speech Separation using Harmonic Selection

'Harmonic selection' is a process of separating speech signals by assuming that they are almost periodic and can be decomposed as a sum of sinusoids [248, 300]. These speech segments are therefore restricted to vowels or vowel-like sounds and, by taking the Fourier transform of the signal over a finite measurement window, it can be seen that the resultant spectrum has well defined resonant peaks at approximate multiples of the fundamental frequency, *i.e.* the pitch [287]. The position of these peaks varies slightly in frequency corresponding to phase changes during phonemes such as vowel sounds, while the magnitude of the peaks varies with changes in formant bandwidth. The property that the instantaneous or short-time spectrum of the two speech signals are effectively non-overlapping in the short-time Fourier domain facilitates signal separation and, therefore, common separation techniques involve the use of variable comb filters.[7] Parsons [287] uses a filtering approach whereby the locations of the harmonics are detected and attributed to either speaker, rather

[7] See [287] for references.

than assuming the harmonics are at exact at multiples of the pitch frequency. This nonlinear approach extracts these peaks, tracks them in position, and reconstructs the signal using a synthetic spectrum. This basic idea has been extended by a number of authors [327], and co-channel speech suppression using sinusoid models for speech has been discussed in detail by Quatieri and Danisewicz [300].

A related approach to harmonic selection is the harmonic magnitude suppression (HMS) technique [149], a form of spectral subtraction in which an estimate of the magnitude spectrum of the noise, or interference, is subtracted from the magnitude spectrum of the noisy signal. The HMS approach estimates the noise spectrum by attempting to estimate the pitch of at least one of the speakers, and then exploits knowledge of the pitch harmonics to separate the two talkers [149, 273]. HMS is used by Lee and Childers [211] to obtain an initial spectral estimate of each talker; then, the cross-entropy of the two talkers is minimised to obtain a better estimate. A recent attempt at co-channel speaker separation by Morgan *et al.* [255] is based on enhancing the harmonic frequencies of the stronger speaker's speech. A review of other work on co-channel[8] speech separation based on the separation of concurrent harmonic sounds can be found in [88].

2.2.4 Speech Separation using Cepstral Filtering

Stubbs and Summerfield [340–342] compare the harmonic selection procedure suggested by Parsons with the cepstral transformation of speech [303]. This transformation is a means of deconvolving the spectral envelope produced by the vocal tract from the vocal excitation. To demonstrate this, consider a single speaker, and assume that, over the measurement window, the *voiced* speech is stationary. The short-time amplitude spectrum, $S(\omega)$, is given by the product of the excitation, $E(\omega)$, and the vocal-tract transfer function, $H(\omega)$:

$$S(\omega) = H(\omega)\, E(\omega) \tag{2.6}$$

For *voiced* speech, $E(\omega)$ is approximately an impulse train, and $H(\omega)$ is a spectral envelope whose gradient has a magnitude much less than the sharp gradients of the impulse train in the excitation sequence, $E(\omega)$. As such, the cepstral transformation maps the spectral envelope, $H(\omega)$, to a region near the origin of the cepstral domain, and maps the harmonic excitation, $E(\omega)$, to a position well separated from

[8] Single channel signal processing is often referred to as 'co-channel' signal processing.

the origin and, therefore, away from the cepstral components due to $H(\omega)$. Since $H(\omega)$ and $E(\omega)$ have been deconvolved in the cepstral domain, they can be manipulated independently. The cepstral transformation is defined as the inverse Fourier transform of the logarithm of the Fourier transform of the speech signal:

$$\begin{aligned} C(\lambda) &= \mathscr{F}^{-1}\{20 \log S(\omega)\} \\ &= \mathscr{F}^{-1}\{20 \log H(\omega) + 20 \log E(\omega)\} \end{aligned} \tag{2.7}$$

where $\mathscr{F}^{-1}$ denotes the inverse Fourier transform. Essentially, deconvolution occurs because the logarithm of $S(\omega)$ flattens out the spectral envelope and so, by taking the inverse Fourier transform, the components of the cepstrum due to the spectral envelope consist of 'low frequencies' whilst, since $\log E(\omega)$ is almost sinusoidal in ω, the component due to $E(\omega)$ is simply an impulse with cepstrum frequency, or *quefrency*, equal to the frequency of the pitch. So, for example, if the pitch peak in the cepstrum is attenuated, the harmonic excitation is reduced, while if the pitch peak is fully attenuated, the voice can be reduced to a whisper. This procedure can be used to attenuate an interfering voice although, since transformation (2.7) is a nonlinear operation, the success of the filtering operation usually requires one voice to be more intense than the interfering voice [340].

The cepstral transformation achieves separation by using several characteristics of the speech signals: first, that the pitch of two speakers usually differs and, second, that the logarithm of the gain of the vocal tract transfer function contains low frequency components. The cepstral transformation is nonlinear and relies on the disjointness of different speech signals in the cepstral domain. It is similar to the harmonic selection technique in that it assumes speech has an impulsive spectrum, but more effectively removes the effect of the vocal tract transfer function. Signal separation in the cepstral domain is a special case of generalised linear filtering [278, 282], [281, Chapter 10] and this is discussed in the next section.

2.2.5 Generalised Linear Filtering (Homomorphic Signal Processing)

Oppenheim *et al.* [278, 282], [281, Chapter 10] discusses a generalisation of linear filtering which allows the theory developed in the following chapters to be applied to systems in which signals have been non-additively combined. This technique, also known as homomorphic signal processing, is based on operators which trans-

form non-additively combined signals into a vector space in which the transformed signals are additively combined. A class of nonlinear filters used on the original signals then correspond to linear transformations on this vector space.

2.2.6 Separation of Periodic Signals

There are many applications where two periodic signals need to be separated, and by far the most typical example is the separation of a mother's electrocardiogram from that of the foetal electrocardiogram. Other applications include the recovery of multiple sinusoids in noise, interference rejection in communication systems which transmit information using narrowband bandpass signals, and in the area of concurrent vowel separation in co-channel speech separation. The separation of periodic signals can be achieved using a comb filter provided the period of the signals are different, so that they do not overlap in the frequency domain. This property has been used in the separation of speech signals as discussed in section §2.2.3.

If mixed signals overlap or are extremely close to one another both in the frequency domain and the time domain, then conventional filtering techniques can prove ineffective due to the large transition band of the comb filter and the problems of resolving overlapping spectral components. Zou and Unbehauen [420] produce a matrix algebraic theory for the separation of potentially spectrally overlapping periodic discrete time signals. The method is based on the fact that if signal $x_1(t)$ has period T_1 and is known for some $t \in \mathscr{T}$, then so is $x_1(t + mT_1)$, $\forall m \in \mathbb{Z}$, and it follows that if $x_1(0)$ is known, periodic signals can be separated given enough samples. Since $x_2(t)$ is also periodic, it can be shown that all the samples of $x_2(t)$ in one period can be calculated using $x_2(t) = x(t) - x_1(0)$ at times $t = nT_1$, $\forall n \in \mathbb{Z}$, where it is assumed that the periods of $x_1(t)$ and $x_2(t)$ are known. Since $x_1(0)$ is unlikely to be known, $x_1(t)$ and $x_2(t)$ can only be recovered to within an additive constant. The matrix algebraic separation and comb filter separation techniques are discussed further by Santhanam and Maragos [315, 316] with particular emphasis on the demodulation of discrete multicomponent AM-FM signals; the algebraic separation approach and comb filters are compared in [316]. An alternative approach to the problem of closely spaced frequencies is the use of the Bayesian methodology [283] which, with suitable prior knowledge, can estimate the distribution of overlapping components between the two constituent signal components.

2.3 Chapter Summary

This chapter has provided a general introduction to the problem of single channel signal separation, and reviews some of the current techniques for the separation of two signals given one sensor. The techniques discussed in the literature are designed for applications ranging from the separation of overlapping speech, to separating electrocardiograms and, as such, their methodology are quite disparate. Although this review is by no means exhaustive, and there exist in the literature many variations of the examples given above, it is clear that no general technique exists to yield a solution to the problem. However, a consistent theme throughout these methods is the way in which the signals are transformed so as to attempt to view the problem from a different angle, leading to new insights into how solutions may be sought; this will be a running theme throughout this work.

3

Linear Time-Varying System Theory

WHILE the analysis of LTI systems and *stationary stochastic processes* is a well developed area [286, 335], the analysis of *linear time-varying* systems and *nonstationary stochastic processes* is less so, although there are some excellent books which have attempted to tackle the subject [84, 296, 339]. This chapter collates relevant nonstationary system theory which leads to the separability criteria discussed in Chapter 4. Section §3.1 reviews the nonstationary autocorrelation function and Wiener-Hopf filter, while section §3.2 reviews the notion of stochastic transforms and nonstationary power spectra, a natural extension to the concept of stationary power spectra. Section §3.3 then investigates the notion of transfer functions for linear time-varying systems.

3.1 Signal Separation using LTV filters

In order to design a LTV filter for the separation of stochastic signals, a criterion must be specified to indicate the choice of filter parameters. This criterion is of-

ten specified in terms of a cost function and, to make the separability problem tractable, a simplification is made by choosing a single fixed cost function. This section describes the cost function and the resulting form of the LTV filter.

3.1.1 Perfect Signal Separation and Signal Separability

'Perfect' signal separation is defined as the recovery of the desired stochastic signal, $\mathbf{d}(t)$, from the corrupted observation, $\mathbf{x}(t)$, as 'accurately' as possible, with the adoption of the following definition:

Definition 2 (Perfect Signal Separation). 'Perfect' separation of deterministic signals is achieved when the estimate of the desired signal identically equals the desired signal for all desired time instances. 'Perfect' separation of stochastic signals is achieved when the mean squared error (MSE), $\sigma^2(t)$, between the estimate of the desired signal, $\hat{\mathbf{d}}(t)$, and the actual desired signal, $\mathbf{d}(t)$,

$$\sigma^2(t) = E\left[\varepsilon^2(t)\right] \triangleq E\left[\left|\hat{\mathbf{d}}(t) - \mathbf{d}(t)\right|^2\right] \tag{3.1}$$

is zero at each desired time instance; $\{\sigma^2(t) = 0; \forall t \in \mathscr{T}\}$. ♢

Bode and Shannon [43] provides a stimulating discussion of the problems and consequences of using MSE as an error criterion. Whilst there are problems with this criterion, it is intellectually and intuitively sound, and will be used as the definition of the accuracy of separation. Clearly, for good separability it is desirable to minimise the MSE. Further discussion of the MSE as an error criterion may be found in, amongst others, the paper by Al-Chalabi [6].

3.1.2 The Autocorrelation Function

As a result of using a second-order cost function, it proves necessary to use the second-order statistics of random signals. The autocorrelation function (ACF) of a stochastic process is a very important statistical property describing the signal in the time domain. The ACF of a nonstationary stochastic process at times $(t, \tau) \in \mathbf{T} = T \times T,\ T \subset \mathbb{R}$ is defined by:[1]

$$R_{xx}(t, \tau) = E_{p(x)}\left[\mathbf{x}(t)\,\mathbf{x}^*(\tau)\right], \quad (t, \tau) \in \mathbf{T} \tag{3.2a}$$

[1] For simplicity, and without loss of generality, all processes are assumed to have zero mean.

where $E_{p(\mathbf{x})}[f(\mathbf{x})]$ denotes the expectation of $f(\mathbf{x})$ w. r. t. the probability density function (pdf) of $\mathbf{x}$. In the case of a second-order or *wide sense stationary* (WSS) process, the ACF is a function of the time difference $(t-\tau) \in \mathbb{R}$ only:

$$R_{xx}(t,\tau) = R_{xx}(t-\tau,0) \triangleq R_{xx}(t-\tau) \tag{3.2b}$$

The problem of estimating the auto- and cross-correlation functions of a nonstationary stochastic process is hindered by the fact that ensemble averages are extremely burdensome or impossible to obtain. Existing approaches [110, 242, 243, 332] to this estimation problem rely on factorisation of the autocorrelation function, as discussed in sections §4.2 and §4.3.

Definition 3 (Correlations for Additive Signals). When an observed signal, $\mathbf{x}(t)$, is composed of a desired signal, $\mathbf{d}(t)$, and a noise component, $\mathbf{n}(t)$, where $\mathbf{d}(t)$ and $\mathbf{n}(t)$ may be correlated, the auto- and cross-correlations are given by:

$$\left.\begin{aligned} R_{xx}(t,\tau) &= R_{dd}(t,\tau) + R_{nn}(t,\tau) \\ &\quad + R_{dn}(t,\tau) + R_{nd}(t,\tau) \\ R_{dx}(t,\tau) &= R_{dd}(t,\tau) + R_{dn}(t,\tau) \end{aligned}\right\}, \quad (t,\tau) \in \mathbf{T} \tag{3.3a}$$

This is the *additive case*. When $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are independent, the cross-correlations are zero and, in this *independent additive case*, the ACFs are given by:

$$\left.\begin{aligned} R_{xx}(t,\tau) &= R_{dd}(t,\tau) + R_{nn}(t,\tau) \\ R_{dx}(t,\tau) &= R_{dd}(t,\tau) \end{aligned}\right\}, \quad (t,\tau) \in \mathbf{T} \tag{3.3b}$$

♢

3.1.3 The Wiener-Hopf Filter

Since separation is achieved by LTV filtration of the observed signal, then

$$\hat{\mathbf{d}}(t) = \int_{\mathscr{T}} h(t,\alpha)\,\mathbf{x}(\alpha)\,d\alpha, \quad \forall t \in T \tag{3.4}$$

where the impulse response of the filter, $h(t,\tau)$, is the response at time t given an impulse occurred at its input at time τ.[2] The cost function dictates that the filter is designed to minimise $\sigma^2(t)$, and this leads to the nonstationary Wiener-Hopf filter (WHF):

[2]The form of this integral for LTV systems is discussed further in section §3.3.

Theorem 1 (Nonstationary Wiener-Hopf Filter) *The filter,* $h(t,\tau)$*, minimising the MSE between* $\hat{\mathbf{d}}(t)$ *and* $\mathbf{d}(t)$ *of equation* (3.4) *is called the Wiener-Hopf filter and is given by the solution of the convolution:*[3]

$$R_{dx}(t,\tau) = \int_{\mathscr{T}} h(t,\alpha)\, R_{xx}(\alpha,\tau)\ d\alpha, \quad \forall (t,\tau) \in T \times \mathscr{T} \tag{3.5a}$$

with the MSE, given by:

$$\sigma^2(t) = R_{dd}(t,t) - \int_{\mathscr{T}} h(t,\alpha)\, R_{dx}(t,\alpha)\ d\alpha, \quad \forall t \in T \tag{3.5b}$$

For 'perfect separation' $\{\sigma^2(t) = 0;\ \forall t \in T\}$ *and, thus, the ACFs must satisfy:*

$$R_{dd}(t,t) = \int_{\mathscr{T}} h(t,\alpha)\, R_{dx}(t,\alpha)\ d\alpha, \quad \forall t \in T \tag{3.5c}$$

The Wiener filter in discrete time can be obtained by replacing the integrals by summations in equations (3.5)*, as shown in Appendix C.2.*

PROOF. There were many contributions in the derivation of the nonstationary Wiener filter, beginning with the original treatise by Wiener in 1942 [394], simplified by Bode and Shannon [43], and then developed by, *inter alios*, Zadeh and Ragazzini [419], Booton [46], Davis [87] and Bendat [33]. The derivation involves differentiating (3.1) w. r. t. $h(t,\tau)$ and setting to zero; for example, see Bendat [33]. A proof is contained in Appendix C.1 for completeness. □

Theorem 2 (Linear Time-Invariant Wiener-Hopf Filter) *If* $\mathbf{d}(t)$ *and* $\mathbf{n}(t)$ *are wide sense stationary, the Wiener-Hopf filter is linear time-invariant.*

PROOF. This well known result follows by substituting equation (3.2b) into (3.5a):

$$R_{dx}(t-\tau) = \int_{\mathbb{R}} h(t,\alpha)\, R_{xx}(\alpha-\tau)\ d\alpha$$

and setting $t \rightarrow t+\hat{\tau}$, $\tau \rightarrow \tau+\hat{\tau}$ and $\hat{\alpha} = \alpha - \hat{\tau}$,

$$R_{dx}(t-\tau) = \int_{\mathbb{R}} h(t+\hat{\tau},\alpha+\hat{\tau})\, R_{xx}(\alpha-\tau)\ d\alpha$$

from which it follows $h(t+\hat{\tau},\tau+\hat{\tau}) = h(t,\tau)$, $\forall \hat{\tau} \in \mathbb{R}$. Thus, $h(t,\tau)$ is LTI. □

[3]The formulation of the Wiener-Hopf filter used here assumes, for clarity, that the input and output processes of the filter in equation (3.4) are real.

3.1.4 A Note on the Kalman Filter

In 1960, Kalman [174] provided an alternative method of formulating the least-squares filtering problem to that of the Wiener filter. The Kalman filter [55, 151, 174, 193] requires identical prior knowledge as required for the Wiener filter and produces the same results. Although the Kalman filter provides a practical technique for linear filtering, the Wiener filter seems to lend itself naturally to analysing the problem of signal separability. Since the solution to the separability problem consists of constraints on the prior knowledge, the results are applicable independently of the implementation of linear filtering.

3.2 Power Spectra for Nonstationary Stochastic Processes

The *autocorrelation function*[4] of a stochastic process is a very important statistical property describing the signal in the time domain, and contains all the prior information required for the formulation of the Wiener-Hopf filter. The *power spectrum* of a stationary stochastic process, $\mathbf{x}(t)$, is a powerfull tool in the analysis of LTI systems, and is defined as the Fourier transform [60, 220, 284] of the ACF given in equation (3.2b):

$$\mathcal{P}_{xx}(\omega) = \int_{\mathsf{T}} \mathrm{R}_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau, \quad \omega \in \Omega \subset \mathbb{R} \tag{3.6}$$

The ACF of a nonstationary stochastic process given by (3.2a) is a well defined two-dimensional deterministic function. Hence, it would be useful to extend the definition of the power spectrum for stationary processes to the nonstationary case.

3.2.1 Notion of Power Spectrum for Nonstationary Signals

The notion of a *power spectrum* for nonstationary signals has been investigated in great depth for the past four decades, beginning with the classic works of Gabor [108], and is usually referred to as *time-frequency analysis*. Cohen, in the classic review paper [75], gives an excellent overview of time-frequency analysis for

[4] See section §3.1.2 for a definition of the ACF.

deterministic signals, notably discussing the Wigner-Ville, Page, and Rihacek [312] distributions.[5] More recently, Hammond and White [141] give an excellent review discussing other time-frequency distributions, for example Loèves' Harmonizable processes [82, 97, 126, 328]. Harmonizable processes are also known as *Wold-Cramer* decompositions and leads to the definition of the *Wold-Cramer* evolutionary spectrum [296, 321]. The Wigner-Ville, Page and Rihacek distributions, among many others, fit into what is known as 'Cohen's class' of bilinear time-frequency distributions defined by the general relation [75]:

$$\mathcal{X}_{\phi(\nu,\tau)}(\lambda, t) = \iiint_{-\infty}^{\infty} \phi(\nu, \tau)\, x\left(\eta + \frac{\tau}{2}\right)\, x\left(\eta - \frac{\tau}{2}\right)\, e^{2\pi j(\nu\eta - \nu t - \lambda\tau)}\, d\nu\, d\eta\, d\tau \qquad (3.7)$$

where $\phi(\nu, \tau)$ is called the *kernel* of the distribution: the Wigner-Ville, Page and Rihacek distributions are obtained by choosing the kernel as $\phi(\nu, \tau) = 1$, $\phi(\nu, \tau) = e^{j|\nu|\tau/2}$ and $\phi(\nu, \tau) = e^{j\nu\tau/2}$, respectively. Cohen's class of bilinear distributions have high resolution, but suffer from cross-terms for multicomponent signals, whereas linear time-frequency transforms, such as the *Wold-Cramer* decomposition, Gabor transform [108, 391], short-time Fourier transform (STFT), and wavelet transform [86], do not have cross-terms for multicomponent signals, but may lack high resolution. Since this thesis mainly considers the separation of a linear combination of signals, only linear transforms are considered, not bilinear forms. A comprehensive and far reaching contribution was made by Mark [240] in which the 'physical' spectrum was introduced; Loynes [229] and Turner [372], *inter alios*, discuss the concept of the power spectrum for nonstationary processes and, in general, the discussion centres around the concept of 'time-varying spectra'. Priestley [295, 296] introduces the notion of the evolutionary spectrum (ES), and numerous researchers have discussed its theoretical framework [229, 355, 368], its applications [1, 2, 83, 139, 367, 369], its estimation [101, 187] and, more recently, its generalisations, for example, [89–91, 196, 246]. In [365], Tsao investigates the form of time-varying covariance functions for nonstationary processes based on Priestley's evolutionary spectrum and, in [366], has investigated the use of the ES to test whether a stochastic process is nonstationary. Lampard [204] generalises the Wiener-Khintchine theorem [60] to nonstationary processes,[6] and that definition of the power spectrum is essentially equivalent to the Page distribution [75].

[5] Note that the Fourier transform of the so-called *local autocorrelation function* $R_{xx}\left(t - \frac{\tau}{2}, t + \frac{\tau}{2}\right)$ is equivalent to the Wigner-Ville distribution for stochastic processes [296].

[6] See [395] for a discussion on the estimation of Lampard's definition of the spectrum from a single realisation of a process.

Cohen [76, 77] reviews a general approach for obtaining joint signal representations using general linear decompositions which analyse signals in terms of physical quantities other than just time and frequency, for example, time and scale in wavelet transforms, and in [318] Sayeed and Jones discuss the connection between current approaches to joint signal representations. Cohen [78] also discusses the generalisation of the stationary autocorrelation function and its power spectral density in terms of a general basis set, giving a corresponding generalisation of the Wiener-Khintchine theorem. Silverman [328] defines a *locally stationary process* through a separable correlation function of the form $R_{xx}(t,\tau) = m\left(\frac{t+\tau}{2}\right) r(t-\tau)$, and generalises the Wiener-Khintchine theorem for such processes. Further investigations of correlation functions and power spectra for nonstationary processes can be found in [11, 97, 188, 213, 241, 372, 392] and references therein, and a general review of time-frequency analysis over the past 50 years has been given by Cohen in [75, 79].

For many of these time-frequency distributions, much emphasis is placed on ensuring there is a meaningful physical interpretation of the power spectra. Each of these power spectral representations have the same motivation behind their development: to find another domain in which the statistical properties of a signal can be represented in a more 'useful' form, or to provide an environment in which a problem becomes easier to solve; this is desirable for the separability problem.

The development of the generalised power spectrum (GPS) in the following sections, of which Priestley's evolutionary spectrum and Loèves' Harmonizable processes are special cases (see section §3.2.5), is based on the theory of linear signal decomposition for representing a stochastic process on an arbitrary power spectral domain. This theory naturally leads to the development of separability criteria for a specific set of signal classes. For purely nonstationary signals, there can exist an ambiguity in the choice of this spectral domain, but often the distinguishing features of signal classes suggest a natural choice, and a separability criterion can be developed assuming the signal is represented in that particular domain. In the stationary case, as shown in section §3.2.4.3, there is no ambiguity in the choice of kernel and, therefore, there arises just a single separability criterion in the case where there is 'limited' prior knowledge.

3.2.2 Spectra of Deterministic Processes

To develop the foundations for definitions of power spectra for stochastic signals, it is instructive to discuss the representation of a deterministic signal on an arbitrary

domain by the use of the general integral transform. This transform effectively converts the basis functions of the signal space to a new set of basis functions. For example, when a time-domain signal is transformed into the Fourier-domain, the basis set $\{\delta(t-\tau),\ (t,\tau) \in T \times T \subset \mathbb{R} \times \mathbb{R}\}$ in the time-domain are changed to the exponential basis set $\{e^{-j\omega t},\ (t,\omega) \in T \times \Omega \subset \mathbb{R} \times \mathbb{R}\}$ in the Fourier domain.

3.2.2.1 Continuous Time and Continuous Spectral Domains

The representation of a *continuous* time deterministic signal, $x(t)$, on an arbitrary *continuous* spectral domain, $\lambda \in \Lambda \subset \mathbb{C}$, is $X(\lambda)$, defined by the general integral transform:

$$X(\lambda) = \int_T x(t)\, K(t,\lambda)\, dt\ ,\ \forall \lambda \in \Lambda \tag{3.8a}$$

where Λ is the region in the signal space in which the representation $X(\lambda)$ lies, and the function $K(t,\lambda)$ is called the *direct transform basis kernel*. Conversely, $X(\lambda)$ may be represented on the time-domain $t \in T \subset \mathbb{R}$ as $x(t)$:

$$x(t) = \int_\Lambda X(\lambda)\, k(\lambda,t)\, d\lambda\ ,\ \forall t \in T \tag{3.8b}$$

where $k(\lambda,t)$ is called the *inverse transform* or *reciprocal basis kernel* of $K(t,\lambda)$. The transformation is assumed isomorphic so, for a given $x(t)$, there exists a unique $X(\lambda)$ and, conversely, for a given $X(\lambda)$, there exists a unique $x(t)$. By extending the definition of a function to include the Dirac δ function [220],[7] possessing the property:

$$f(t) = \begin{cases} \int_T \delta(t-\tau) f(\tau)\, d\tau & ,\quad \forall t \in T \\ 0 & ,\quad \forall t \notin T \end{cases} \tag{3.9}$$

it can be shown that the kernels $k(\lambda,t)$ and $K(t,\lambda)$ must satisfy:

$$\delta(t-\tau) = \int_\Lambda K(t,\lambda)\, k(\lambda,\tau)\, d\lambda\ ,\ \forall (t,\tau) \in \mathbf{T} \tag{3.10a}$$

$$\delta(\lambda-\hat{\lambda}) = \int_T k(\lambda,t)\, K(t,\hat{\lambda})\, dt\ ,\ \forall (\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda} \tag{3.10b}$$

[7] The Dirac delta function is a generalised function, and is used with the understanding that, formally, the theory can be rewritten in terms of generalised functions to produce rigorous results.

where $\boldsymbol{\Lambda} = \Lambda \times \Lambda$ and $\mathbf{T} = \mathsf{T} \times \mathsf{T}$. Any pair of such functions satisfying these equations for $(t, \lambda) \in \mathsf{T} \times \Lambda$ is called a transform pair over $\mathsf{T} \times \Lambda$. If $k(\lambda, t) = K^*(t, \lambda)$, they are said to be self-reciprocal kernels, and correspond to *orthonormal* basis in the finite dimensional case. A useful property of all orthonormal transforms is energy conservation:

Theorem 3 (Parseval's Energy Conservation Property) *If the basis functions of the transformation defined by* (3.8) *are self-reciprocal such that* $k(\lambda, t) = K^*(t, \lambda)$, *where* $\lambda \in \Lambda \subset \mathbb{R}$, *then* (3.8) *possesses Parseval's energy conservation property:*

$$\int_{\mathsf{T} \subset \mathbb{R}} |x(t)|^2 \, dt = \int_{\Lambda \subset \mathbb{R}} |X(\lambda)|^2 \, d\lambda \tag{3.11}$$

PROOF. The proof follows directly using the expansions of equations (3.8) with $K(t, \lambda) = k^*(\lambda, t)$ and by writing $|x(t)|^2 = x(t)\, x^*(t)$, $|X(\lambda)|^2 = X(\lambda)\, X^*(\lambda)$:

$$\begin{aligned} \int_{\mathsf{T}} |x(t)|^2 \, dt &= \int_{\mathsf{T}} x(t) \int_{\Lambda} X^*(\lambda)\, k^*(\lambda, t) \, d\lambda \, dt \\ &= \int_{\Lambda} X^*(\lambda) \int_{\mathsf{T}} x(t)\, k^*(\lambda, t) \, dt \, d\lambda = \int_{\Lambda} |X(\lambda)|^2 \, d\lambda \end{aligned} \tag{3.12}$$

□

Spectral decompositions of deterministic processes in the form of (3.8) have been discussed by numerous authors in the field over the past fifty years, including [17, 31, 84, 150, 413–415, 417, 418]. Examples include series expansions in Bessel functions, prolate spherical functions, orthogonal polynomials, and systems of orthogonal functions defined by linear differential or difference equations (see section §3.3.5 and [17, 84, 150]); Naylor investigates the decomposition of discrete LTV systems based on its singular values [272], and Harmuth [150] discusses the concept of 'generalised frequency' for some of these expansions, notable the Walsh-Fourier transform.

The use of the delta-function may be avoided by writing all integrals, such as (3.8b), in the Stieltjes integral form [60, 220]:

$$x(t) = \int_{\Lambda} k(\lambda, t) \, dX(\lambda) \;, \; \forall t \in \mathsf{T} \tag{3.13}$$

where, if $X(\lambda)$ is absolutely continuous w. r. t. to the Lebesgue measure, $dX(\lambda) \equiv X(\lambda)\, d\lambda$ [60]. However, Stieltjes integrals can sometimes obscure the ideas behind the results presented in this work, as well as the parallels between continuous and discrete time cases. Therefore, integrals such as (3.8) are used with the understand-

ing that generalised functions and, or, Stieltjes integrals can be employed to produce formally rigorous results. As such, for brevity, all integrals are assumed to exist for all functions which arise in this work.

3.2.2.2 Discrete Time and Discrete Spectral Domains

The spectral decomposition in the finite *discrete* time, finite *discrete* spectral case is conceptually simpler than the continuous case, since a transform from one domain to another can be viewed as a change in basis vectors.[8] The representation of a discrete time deterministic signal $\{x(n),\, n \in \mathscr{N} \subset \mathbb{Z}\}$ on an arbitrary finite support discrete spectral domain is $\{X(p),\, p \in \mathscr{P} \subset \mathbb{Z}\}$, and are related by

$$X(p) = \sum_{n \in \mathscr{N}} x(n)\, K(n,p), \quad x(n) = \sum_{p \in \mathscr{P}} X(p)\, k(p,n) \tag{3.14}$$

corresponding to (3.8) for the continuous time–spectral case. The support of each domain is finite, for example, $\mathscr{N} = \{0,\ldots,N-1\}$ and $\mathscr{P} = \{0,\ldots,P-1\}$. Further, to avoid *degenerate* or *redundant* cases, it is usually assumed that $N = P$. The representations on each domain must be uniquely related so, substituting one transform equation into the other gives rise to the kernel relationships:

$$\sum_{p \in \mathscr{P}} K(n,p)\, k(p,\hat{n}) = \delta(n-\hat{n}) \;, \quad (n,\hat{n}) \in \mathscr{N} \times \mathscr{N} \tag{3.15a}$$

$$\sum_{n \in \mathscr{N}} k(p,n)\, K(n,\hat{p}) = \delta(p-\hat{p}) \;, \quad (p,\hat{p}) \in \mathscr{P} \times \mathscr{P} \tag{3.15b}$$

where $\delta(p) = 1$ if $p = 0$ and 0 otherwise. The finite discrete case may be written as a matrix formulation, emphasising the fact that the transform is simply equivalent to a change of basis. Defining the signal and spectral vectors $[\mathbf{x}]_n = x(n)$, $[\mathbf{X}]_n = X(n)$, $n \in \mathscr{N}$, and the kernel matrices $[k]_{pn} = k(p,n)$, $[K]_{np} = K(n,p)$, $(n,p) \in \mathscr{N} \times \mathscr{P}$, then (3.14) can be written compactly as:

$$\mathbf{X} = K^{T}\mathbf{x} \qquad \text{and} \qquad \mathbf{x} = k^{T}\mathbf{X} \tag{3.16}$$

where T denotes matrix transpose $[A^{T}]_{np} = [A]_{pn}$. Furthermore,

$$k\,K = K\,k = I_N \tag{3.17}$$

[8] Often results for the discrete time case can be obtained from their continuous time counterparts by replacing integrals with summations, *mutatis mutandis*.

where $I_N \in \mathbb{R}^{N\times N}$ is the identity matrix, and (3.17) is equivalent to (3.15) written in matrix form. Observe that (3.17) implies the basis kernels must have full rank.

3.2.2.3 Discrete Time and Continuous Spectral Domains

The case of discrete time signals of infinite duration leading to a continuous spectrum, and of finite, or periodic, continuous time signals leading to a discrete spectrum, are considered, *mutatis mutandis*, as special cases of the continuous-time – continuous-spectral case in some appropriate limit.

3.2.3 Stochastic Spectral Transforms

In LTI signal theory, spectra are often associated with Fourier transforms and, for deterministic signals, this idea was extended in section §3.2.2 using a general integral transform to represent a function as the superposition or integral of some given basis. This concept may be extended to random signals, with a particular realisation of a stochastic process decomposed as the superposition or integral of basis functions with stochastic coefficients. This is a generalisation of both the Fourier transform of a stochastic process, as discussed in [286, 335], and the Karhunen-Loève (KL) transform [286, 335, 353] (see Appendix B.1) – another widely known integral transform. Note that when the stochastic processes are stationary, the Fourier transform is a special case of the KL transform (see Theorem 23 in Appendix B.1). In the KL transform, the kernel $\{k(t,\lambda)\}$ is self-reciprocal and, as such, is a set of orthonormal functions. A useful characteristic of the Karhunen-Loève and Fourier transforms is that the resulting spectral components, $\mathbf{X}(\lambda)$, are uncorrelated.

3.2.3.1 Continuous Time and Continuous Spectral Domains

The stochastic spectral representation of a *continuous* stochastic process, $\mathbf{x}(t)$, on an arbitrary *continuous* spectral domain, $\lambda \in \Lambda \subset \mathbb{C}$, is $\mathbf{X}(\lambda)$, defined by:

$$\mathbf{X}(\lambda) = \int_T \mathbf{x}(t)\, K(t,\lambda)\, dt\,, \qquad \forall \lambda \in \Lambda \tag{3.18a}$$

where this integral is interpreted in a mean square (MS) limit. Conversely, $\mathbf{X}(\lambda)$ may be represented on the time domain, $t \in \mathsf{T} \subset \mathbb{R}$, as $\hat{\mathbf{x}}(t)$:

$$\hat{\mathbf{x}}(t) = \int_{\Lambda} \mathbf{X}(\lambda)\, k(\lambda, t)\, d\lambda\,, \qquad \forall t \in \mathsf{T} \tag{3.18b}$$

Theorem 4 (Mean Square Equality of Integral Transforms) *The representation* $\hat{\mathbf{x}}(t)$ *equals* $\mathbf{x}(t)$ *in the MS sense:*

$$\mathrm{E}\left[|\mathbf{x}(t) - \hat{\mathbf{x}}(t)|^2\right] = 0\,, \quad \forall t \in \mathsf{T} \tag{3.19}$$

The proof of this results is given with the proof of Theorem 6 on page 48. It is of interest to look at the parallels of this decomposition with the innovations representation of a stochastic process.

3.2.3.2 Innovations Representation of a Stochastic Process

Theorem 5 (Innovations of a Random Process) *Any stochastic process,* $\mathbf{x}(t)$*, satisfying certain weak conditions as detailed in Cramér [82, pp. 70], can be expressed in the form:*

$$\mathbf{x}(t) = \int_{\mathsf{T}} g(t, \tau)\, \mathbf{z}(\tau)\, d\tau, \quad \forall t \in \mathsf{T} \tag{3.20}$$

where $\mathbf{z}(\tau)$ *is an orthogonal stationary stochastic process, and* $g(t, \tau)$ *is a deterministic function. This representation may be regarded as passing a stationary white noise process through a LTV filter with impulse response* $g(t, \tau)$.

PROOF. See, *e.g.*, Cramér [82], or [285,286]. □

The innovations expansion is equivalent to the spectral representation (3.18b) and, as such, is not unique. They can be shown to be equivalent by writing the stationary process, $\mathbf{z}(t)$, using the Fourier transform:

$$\mathbf{z}(t) = \frac{1}{2\pi} \int_{\Omega} \mathbf{Z}(\omega)\, e^{-j\omega t}\, d\omega\,, \quad \text{where } \Omega \subset \mathbb{R}$$

Thus, from equation (3.20):

$$\mathbf{x}(t) = \frac{1}{2\pi} \iint_{\mathsf{T}\Omega} g(t, \tau)\, \mathbf{Z}(\omega)\, e^{-j\omega\tau}\, d\omega\, d\tau \equiv \int_{\Omega} G(t, \omega)\, \mathbf{Z}(\omega)\, d\omega \tag{3.21}$$

which can be rewritten in the form of (3.18b), where $G(t,\omega)$ is the Fourier transform w. r. t. to τ of $g(t,\tau)$ with t considered as a parameter.

3.2.3.3 Generalised Innovations of a Stochastic Process

Suppose that equation (3.18b) is written in the form of equation (3.21); this implies:

$$E\left[\hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^*(\tau)\right] = \iint_{\Omega\Omega} G(t,\omega)\,G^*(\tau,\hat{\omega})E\left[\mathbf{Z}(\omega)\,\mathbf{Z}^*(\hat{\omega})\right]\,d\omega\,d\hat{\omega}\,,\quad \forall(t,\tau)\in\mathbf{T}$$

or, since $\mathbf{Z}(\omega)$ is a stationary process and, therefore, $E\left[\mathbf{Z}(\omega)\,\mathbf{Z}^*(\hat{\omega})\right] \equiv S(\omega)\,\delta(\omega-\hat{\omega})$ (see [286]), this simplifies to:

$$R_{\hat{x}\hat{x}}\,(t,\tau) = \int_{\Omega} G(t,\omega)\,G^*(\tau,\omega)\,S(\omega)\,d\omega,\quad \forall(t,\tau)\in\mathbf{T} \tag{3.22}$$

Such a definition for a power spectrum, where the spectrum is a one-dimensional function, is slightly restrictive and arises because of the assumption that the process must be generated by passing stationary white noise through a LTV filter. The KL transform is an example of a representation in the form of equation (3.22). However, to increase the set of possible domains on which the ACF may be represented, it may be assumed that a form of (3.18b) exists in which $\mathbf{X}(\lambda)$ is not necessarily statistically orthogonal. This spectral transform leads naturally to the definition of a two-dimensional (2-D) power spectrum introduced in section §3.2.4.

3.2.3.4 Autocorrelation of the Stochastic Integral Representation

The spectral decomposition of equation (3.18a) implies the ACF of the random variable $\mathbf{X}(\lambda)$, $R_{XX}(\lambda,\hat{\lambda}) \triangleq E[\mathbf{X}(\lambda)\,\mathbf{X}^*(\hat{\lambda})]$, may be written $\forall(\lambda,\hat{\lambda})\in\boldsymbol{\Lambda}$ as:

$$\mathcal{P}_{\hat{x}\hat{x}}(\lambda,\hat{\lambda}) \triangleq R_{XX}(\lambda,\hat{\lambda}) = \iint_{\mathbf{T}} R_{xx}\,(t,\tau)\;K(t,\lambda)\,K^*(\tau,\hat{\lambda})\,dt\,d\tau \tag{3.23a}$$

Further, (3.18b) implies the ACF of $\hat{\mathbf{x}}(t)$, $R_{\hat{x}\hat{x}}\,(t,\tau)$, may be written $\forall(t,\tau)\in\mathbf{T}$ as:

$$R_{\hat{x}\hat{x}}\,(t,\tau) \equiv E\left[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^*(\tau)\right] = \iint_{\boldsymbol{\Lambda}} \mathcal{P}_{\hat{x}\hat{x}}(\lambda,\hat{\lambda})\,k(\lambda,t)\,k^*(\hat{\lambda},\tau)\,d\lambda\,d\hat{\lambda} \tag{3.23b}$$

3.2.4 Generalised Power Spectrum

As discussed in section §3.2 the ACF of a nonstationary stochastic process, $R_{xx}(t,\tau)$, is a well defined 2-D deterministic function and, therefore, as a natural extension to the concept of the stationary power spectrum [286,335], it may be expressed on an arbitrary spectral domain using a 2-D integral transform:

$$\mathcal{P}_{xx}(\lambda,\hat{\lambda}) = \iint_{\mathbf{T}} R_{xx}(t,\tau)\,\hat{K}(t,\tau,\lambda,\hat{\lambda})\,dt\,d\tau\,, \quad \forall(\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda}$$

Considering the form of the autocorrelation functions of $\mathbf{X}(\lambda)$ and $\hat{\mathbf{x}}(t)$ in equations (3.23), and the innovations representation discussed in section §3.2.3.2, it is natural to assume the kernel $\hat{K}(t,\tau,\lambda,\hat{\lambda})$ admits a separable form: *i.e.* $\hat{K}(t,\tau,\lambda,\hat{\lambda}) \triangleq K(t,\lambda)\,K(\tau,\hat{\lambda})$, $(\lambda,\hat{\lambda},t,\tau) \in \boldsymbol{\Lambda}\times\mathbf{T}$. With such a separable kernel, the following definitions may be made.

3.2.4.1 Autocorrelation Decomposition and Power Spectrum

Definition 4 (Generalised Power Spectrum). The generalised power spectrum (GPS) of the process $\mathbf{x}(t)$ and its inverse relationship are defined as:

$$\mathcal{P}_{xx}(\lambda,\hat{\lambda}) = \iint_{\mathbf{T}} R_{xx}(t,\tau)\,K(t,\lambda)\,K^*(\tau,\hat{\lambda})\,dt\,d\tau\,, \quad \forall(\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda} \tag{3.24a}$$

$$R_{xx}(t,\tau) = \iint_{\boldsymbol{\Lambda}} \mathcal{P}_{xx}(\lambda,\hat{\lambda})\,k(\lambda,t)\,k^*(\hat{\lambda},\tau)\,d\lambda\,d\hat{\lambda}\,, \quad \forall(t,\tau) \in \mathbf{T} \tag{3.24b}$$

where $K(t,\lambda)$ and $k(\lambda,t)$ are related by (3.10). The choice of these kernel functions is not necessarily arbitrary, and an *important* example is given in section §3.2.4.3. ♢

Theorem 6 (Basic GPS Relationships) *Some basic relations follow from the definition of the generalised power spectrum:*

$$\text{(i)}\ \mathcal{P}^*_{xx}(\lambda,\hat{\lambda}) = \mathcal{P}_{xx}(\hat{\lambda},\lambda),\ \text{(ii)}\ \mathcal{P}_{\hat{x}\hat{x}}(\lambda,\hat{\lambda}) = \mathcal{P}_{xx}(\lambda,\hat{\lambda}),\ \text{(iii)}\ R_{\hat{x}\hat{x}}(t,\tau) = R_{xx}(t,\tau)$$

PROOF. Relation (i) follows by conjugating each side of equation (3.24a), and noting $R^*_{xx}(t,\tau) = R_{xx}(\tau,t)$. Relation (ii) follows by comparing (3.23a) with (3.24a);

relation (iii) follows by substituting (ii) into (3.24b) and comparing with (3.23b). □

PROOF (PROOF OF THEOREM 4). Using relation (iii) of Theorem 6,

$$\mathrm{E}\left[|\hat{\mathbf{x}}(t)|^2\right] = \mathrm{R}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}(t,t) \equiv \mathrm{R}_{\mathbf{x}\mathbf{x}}(t,t) = \mathrm{E}\left[|\mathbf{x}(t)|^2\right] \tag{3.25a}$$

Next, noting the product

$$\mathrm{E}\left[\mathbf{X}(\lambda)\,\mathbf{x}^*(t)\right] = \int_{\mathsf{T}} \mathrm{E}\left[\mathbf{x}(\alpha)\,\mathbf{x}^*(t)\right] \mathrm{K}(\alpha,\lambda)\,\mathrm{d}\alpha = \int_{\mathsf{T}} \mathrm{R}_{\mathbf{x}\mathbf{x}}(\alpha,t)\,\mathrm{K}(\alpha,\lambda)\,\mathrm{d}\alpha \tag{3.25b}$$

then using equation (3.18a), and substituting equation (3.25b), yields:

$$\begin{aligned}\mathrm{E}\left[\hat{\mathbf{x}}(t)\,\mathbf{x}^*(t)\right] &= \int_{\Lambda} \mathrm{E}\left[\mathbf{X}(\lambda)\,\mathbf{x}^*(t)\right]\,\mathrm{k}(\lambda,t)\,\mathrm{d}\lambda \\ &= \int_{\mathsf{T}} \mathrm{R}_{\mathbf{x}\mathbf{x}}(\alpha,t) \int_{\Lambda} \mathrm{K}(\alpha,\lambda)\,\mathrm{k}(\lambda,t)\,\mathrm{d}\lambda\,\mathrm{d}\alpha = \mathrm{R}_{\mathbf{x}\mathbf{x}}(t,t)\end{aligned}$$

from using equation (3.10a). Hence,

$$\begin{aligned}\mathrm{E}\left[\hat{\mathbf{x}}(t)\,\mathbf{x}^*(t)\right] &= \mathrm{R}_{\mathbf{x}\mathbf{x}}(t,t) = \mathrm{R}^*_{\mathbf{x}\mathbf{x}}(t,t) \\ &= \mathrm{E}\left[\mathbf{x}(t)\,\mathbf{x}^*(t)\right] = \mathrm{E}\left[\hat{\mathbf{m}}\mathbf{x}^*(t)\,\mathbf{x}(t)\right]\end{aligned} \tag{3.25c}$$

Hence, using (3.25a) and (3.25c), (3.19) readily follows upon expansion. □

3.2.4.2 *Cross-correlation Decomposition and Cross-Power Spectrum*

Suppose that a signal $\mathbf{y}(t)$ admits a similar representation to equations (3.18):

$$\begin{aligned}\mathbf{Y}(\lambda) &= \int_{\mathsf{T}} \mathbf{y}(t)\,\mathrm{K}(t,\lambda)\,\mathrm{d}t\,, \quad \forall\lambda \in \hat{\Lambda}_{\mathbf{Y}} \\ \hat{\mathbf{y}}(t) &= \int_{\hat{\Lambda}_{\mathbf{Y}}} \mathbf{Y}(\lambda)\,\mathrm{k}(\lambda,t)\,\mathrm{d}\lambda\,, \quad \forall t \in \mathsf{T}\end{aligned} \tag{3.26}$$

then the cross-correlation $\mathrm{R}_{\hat{\mathbf{y}}\hat{\mathbf{x}}}(t,\tau) = \mathrm{E}\left[\hat{\mathbf{y}}(t)\,\hat{\mathbf{x}}^*(\tau)\right]$ may be written in the form:

$$\mathrm{R}_{\hat{\mathbf{y}}\hat{\mathbf{x}}}(t,\tau) = \iint_{\hat{\Lambda}_{\mathbf{X}}\hat{\Lambda}_{\mathbf{Y}}} \mathcal{P}_{\hat{\mathbf{y}}\,\hat{\mathbf{x}}}(\lambda,\hat{\lambda})\,\mathrm{k}(\lambda,t)\,\mathrm{k}^*(\hat{\lambda},\tau)\,\mathrm{d}\lambda\,\mathrm{d}\hat{\lambda}\,, \quad \forall(t,\tau) \in \mathbf{T} \tag{3.27}$$

where $\hat{\Lambda}_X \equiv \Lambda_X \oplus \Lambda_0$, $\hat{\Lambda}_Y \equiv \Lambda_Y \oplus \Lambda_0$ and $\mathcal{P}_{\hat{y}\hat{x}}(\lambda, \hat{\lambda}) = E[\mathbf{Y}(\lambda)\,\mathbf{X}^*(\hat{\lambda})]$.[9] Here, $\Lambda_X \subset \Lambda$ and $\Lambda_Y \subset \Lambda$ are the regions over which the spectral components of $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{y}}(t)$, respectively, do *not* overlap, and $\Lambda_0 \subset \Lambda$ is the region over which spectral components of $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{y}}(t)$ *do* overlap. Hence, $\Lambda_X \cap \Lambda_Y = \{\emptyset\}$, $\Lambda_X \cap \Lambda_0 = \{\emptyset\}$, $\Lambda_Y \cap \Lambda_0 = \{\emptyset\}$, and $\Lambda_X \oplus \Lambda_Y \oplus \Lambda_0 \subset \Lambda$. The limits of the integral in (3.27) become:

$$\iint_{\hat{\Lambda}_X \hat{\Lambda}_Y} \equiv \iint_{\Lambda_X \Lambda_0} + \iint_{\Lambda_X \Lambda_Y} + \iint_{\Lambda_0 \Lambda_0} + \iint_{\Lambda_0 \Lambda_Y} \tag{3.28}$$

and it is seen that only the third integral in the RHS of (3.28) is non-zero, since $\mathcal{P}_{\hat{y}\hat{x}}(\lambda, \hat{\lambda}) = E[\mathbf{Y}(\lambda)\,\mathbf{X}^*(\hat{\lambda})] = 0$ if $(\lambda, \hat{\lambda}) \in \Lambda_X \times \Lambda_Y$, and so forth. Thus, using a similar analysis as in section §3.2.4, the generalised cross power spectrum (GCPS) is defined as in Definition 5:

Definition 5 (Generalised Cross-Power Spectrum). The generalised cross power spectrum (GCPS) of the processes $\mathbf{y}(t)$ and $\mathbf{x}(t)$, $\forall(\lambda, \hat{\lambda}) \in \boldsymbol{\Lambda}_0 = \Lambda_0 \times \Lambda_0$, $\Lambda_0 \subset \Lambda$, and its inverse relationship are defined as:

$$\mathcal{P}_{yx}(\lambda, \hat{\lambda}) = \iint_{\mathbf{T}} R_{yx}(t, \tau)\, K(t, \lambda)\, K^*(\tau, \hat{\lambda})\, dt\, d\tau, \quad \forall(\lambda, \hat{\lambda}) \in \boldsymbol{\Lambda}_0 \tag{3.29a}$$

$$R_{yx}(t, \tau) = \iint_{\boldsymbol{\Lambda}_0} \mathcal{P}_{yx}(\lambda, \hat{\lambda})\, k(\lambda, t)\, k^*(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda}, \quad \forall(t, \tau) \in \mathbf{T} \tag{3.29b}$$

where $K(t, \lambda)$ and $k(\lambda, t)$ are related by equations (3.10), and Λ_0 is the region over which the stochastic spectra of $\mathbf{x}(t)$ and $\mathbf{y}(t)$ overlap. The definition of the GCPS gives the relation $\mathcal{P}^*_{yx}(\lambda, \hat{\lambda}) = \mathcal{P}_{xy}(\hat{\lambda}, \lambda)$. ♢

3.2.4.3 Power Spectra for Stationary Processes

Theorem 7 (Power Spectral Density for Stationary Processes) *The ACF of a WSS process admits representation in the time domain and Fourier domain [286, 335, 353]. The converse is not true: any general nonstationary stochastic signal which admits representation in the Fourier domain is not necessarily WSS.*

PROOF. The autocorrelation function of a stochastic process $\mathbf{x}(t)$ may be expressed

[9] The *orthogonal direct sum* is denoted by $\oplus$, and defined such that if $A \oplus B = C$, then $A \cup B = C$ and $A \cap B = \{\emptyset\}$.

using equation (3.24b) as:

$$R_{xx}(t,\tau) = \iint_{\Omega\Omega} \mathcal{P}_{xx}(\omega,\hat{\omega})\, k(\omega,t)\, k^*(\hat{\omega},\tau)\, d\omega\, d\hat{\omega},\ \forall (t,\tau) \in \mathbf{T} \tag{3.30a}$$

where $\Omega \subset \mathbb{C}$. Since $R_{xx}(t,\tau) = R_{xx}(t-t_0,\tau-t_0)$, $\forall t_0 \in \mathbb{R}$ for a WSS process then, setting $t_0 = \tau$ and using Definition 4, it follows:

$$\begin{aligned} R_{xx}(t,\tau) &= R_{xx}(t-\tau,0) \equiv R_{xx}(t-\tau) \\ &= \iint_{\Omega\Omega} \mathcal{P}_{xx}(\omega,\hat{\omega})\, k(\omega,t-\tau)\, k^*(\hat{\omega},0)\, d\omega\, d\hat{\omega} \end{aligned} \tag{3.30b}$$

which may be written as:

$$R_{xx}(t) = \int_{\Omega} \mathcal{P}_{xx}(\omega)\, k(\omega,t)\, d\omega \text{ where } \mathcal{P}_{xx}(\omega) = \int_{\Omega} \mathcal{P}_{xx}(\omega,\hat{\omega})\, k^*(\hat{\omega},0)\, d\hat{\omega} \tag{3.30c}$$

Equation (3.30c) may be used as a foundation for defining the 'equivalent stationary' power spectral density (PSD) of a nonstationary process. Comparing equations (3.30a) and (3.30b), it follows that:

$$\mathcal{P}_{xx}(\omega,\hat{\omega})\,[k(\omega,t)\, k^*(\hat{\omega},\tau) - k(\omega,t-\tau)\, k^*(\hat{\omega},0)] = 0,\ \forall (t,\tau,\omega,\hat{\omega}) \in \mathbf{T} \times \boldsymbol{\Omega} \tag{3.31a}$$

where $\boldsymbol{\Omega} = \Omega \times \Omega$. Hence, either $\mathcal{P}_{xx}(\omega,\hat{\omega}) = 0$, or

$$k(\omega,t)\, k^*(\hat{\omega},\tau) = k(\omega,t-\tau)\, k^*(\hat{\omega},0) \tag{3.31b}$$

Since ω, $\hat{\omega}$, t, and τ are independent variables, equation (3.31b) can only be valid when $\omega = \hat{\omega}$. Thus, the kernel function must satisfy

$$k(\omega,t)\, k^*(\omega,\tau) = k(\omega,t-\tau)\, k^*(\omega,0) \tag{3.31c}$$

The family of functions satisfying this criterion, and which have a unique reciprocal basis, are of the form $k(\omega,t) = \frac{1}{2\pi} e^{jg(\omega)t}$, where $g(\omega)$ is a real function such that $g(\omega) : \mathbb{R} \to \mathbb{R}$. A change of variable in the integral transform of $\tilde{\omega} = g(\omega)$ thus leads to the Fourier kernel. Moreover, notice that since equation (3.31b) can only be true if $\omega = \hat{\omega}$, then it follows from equation (3.31a) that $\mathcal{P}_{xx}(\omega,\hat{\omega}) = 0$, $\forall \omega \neq$

$\hat{\omega}$, as expected. As a result of the positive definiteness of the ACF [286,335], $\mathcal{P}_{xx}(\omega)$ is always positive. □

For deterministic signals, Claasen and Mecklenbräuker [73] define a *stationary signal* as, heuristically speaking, one whose spectral content does not vary with time.

3.2.5 Related Time-Varying Spectra to the GPS

The generalised power spectrum covers, as special cases, a couple of very important power spectral representations that have played an important rôle in signal analysis: the *evolutionary spectrum* and *Harmonizable processes*.

3.2.5.1 The Evolutionary Spectrum

Priestley's evolutionary spectrum [295,296] assumes a process, $\mathbf{x}(t)$, admits representation in the form of equation (3.21), where $\mathbf{Z}(\lambda)$ is an orthogonal stochastic process, such that $E[\mathbf{Z}(\omega)\,\mathbf{Z}^*(\hat{\omega})] = \mathcal{S}_{zz}(\omega)\,\delta(\omega-\hat{\omega})$, $\forall(\omega,\hat{\omega}) \in \Omega\times\Omega \subset \mathbb{R}\times\mathbb{R}$. Furthermore, to attribute some physical meaning to $G(\omega,t)$, it may be written as an amplitude modulated exponential function:

$$G(\lambda,t) \triangleq k(\lambda,t)\,e^{j\lambda t} \tag{3.32}$$

In such a case, the autocorrelation function of equation (3.24b) becomes, using the Stieltjes integral representation:

$$R_{xx}(t,\tau) = \int_{\Lambda} k(\lambda,t)\,k^*(\lambda,\tau)\,e^{j\lambda(t-\tau)}\,d\mathcal{S}_{xx}(\lambda) \tag{3.33}$$

The time-varying spectral density, termed the evolutionary spectrum, is given by:

$$\mathcal{P}_{xx}(\lambda,t) = |k(\lambda,t)|^2\,\mathcal{S}_{xx}(\lambda) \tag{3.34}$$

The class of processes which admit this representation are termed 'oscillatory processes', and an essential requirement by Priestley is that $k(\lambda,t)$ be 'slowly varying' as a function of t for each λ, compared to the exponential function $e^{j\lambda t}$. Priestley and Tong [296,297] have extended this definition to cross-power spectral expansions.

3.2.5.2 Harmonizable Processes

A nonstationary stochastic process, $\mathbf{x}(t)$, is called *Harmonizable* [82, 126] if there exists a second-order spectral decomposition of $\mathbf{x}(t)$, denoted $\mathbf{X}(\lambda)$, given by the Fourier transform relationships. The square modulus of this expansion is often called the Wold-Cramer evolutionary spectrum and, under certain conditions, coincides with Priestley's definition of the evolutionary spectrum [295, 296, 321]. Using the notation in sections §3.2.3 and §3.2.4, $\mathcal{P}_{xx}(\lambda, \hat{\lambda})$ is called Loève's 'generalised power spectral density', where the transform kernels are given by $k(\lambda, t) = \frac{1}{2\pi} e^{j\lambda t}$ and $K(t, \lambda) = e^{-j\lambda t}$. Grace [126] discusses relationships between Harmonizable processes and other definitions of power spectral densities, such as the instantaneous power spectrum (IPS) (Page's distribution). Harmonizable processes are useful when the only knowledge regarding a nonstationary stochastic process is that it is bandlimited. The definition of a Harmonizable process arises naturally from the generalised power spectrum (GPS) as defined in section §3.2.4, and is given by the symplectic 2-D Fourier transform of the ACF.

3.3 Transfer functions

If the separation of stochastic signals is to be achieved using linear time-varying filters, it is important to investigate further the structure of such filters. This general structure is developed in terms of the transform kernels used to represent the signals which the filter operates upon. It is, of course, possible to develop the structure of a linear time-varying filter with respect to any basis set but, as will be seen in section §4.2.2, it is convenient to use a consistent basis set throughout the analysis. In signal theory, the concept of a transfer function of a linear time-invariant system is very important, as filtration becomes multiplication in the spectral domain. It is, thus, natural to extend the definition of a transfer function to LTV systems, and this concept is also helpful when introducing the *ideal filter* in section §4.1. The theory developed is equally valid when the system operates either on deterministic signals, or on stochastic signals.[10]

The approach used here takes advantage of a generalisation of the *bifrequency transfer function* of a filter [32, 73, 107, 409–411], which is of fundamental im-

[10] See section §4.1.2 for a discussion on the filtration of random signals.

portance in the study of LTV systems. The generalised bifrequency transfer function (GBTF) is sometimes called the *kernel function* of the system [347]. An alternative definition of a transfer function, called the *general system function*, of which the well-known Zadeh's transform is a special case [411], is not pursued here. Nevertheless, a direct relationship between these two definitions is outlined in section §3.3.3. References to other work in the literature discussing generalised frequency response concepts for linear time-varying systems can be found in, for example, [21, 84, 417]. For reference, Tong [362] and Tsao [367] consider time-varying filtration of signals which admit evolutionary spectral representations, whilst the concept of filtering in the mixed *time-frequency* domain has been considered by Saleh and Subotic [314] for STFT and Wigner representations, by Portnoff for STFT representations [293], and in Gabor space by Raz *et al.* [103, 104, 313]; further references of time-frequency filtration can be found in [75].

The input–output relation for a linear time-varying system may be specified by the well known time-domain *superposition integral* [84]:

$$\mathbf{y}(t) = \int_{\mathsf{T}} h(t,\tau)\,\mathbf{x}(\tau)\,d\tau \triangleq h(t,\tau) \star \mathbf{x}(\tau), \quad \forall t \in \mathsf{T} \tag{3.35}$$

where $\mathbf{y}(t)$ is the response of the linear system to the input $\mathbf{x}(t)$, and $\star$ denotes convolution over the variable τ as defined in equation (3.35). The two-dimensional function, $h(t,\tau)$, is the LTV impulse response at time t to an impulse occurring at the system input at time τ. The term *impulse response* is not uniquely defined in the LTV system literature and, in many publications, the impulse response is also given by $\hat{h}(t,\tau) = h(t, t-\tau)$, such that the input–output relation is given as:

$$\mathbf{y}(t) = \int_{\mathsf{T}} \hat{h}(t, t-\tau)\,\mathbf{x}(\tau)\,d\tau, \quad \forall t \in \mathsf{T} \tag{3.36}$$

The form in equation (3.35) is used throughout this work since it corresponds to the usual definition for the kernel of a linear integral operator as discussed in section §3.2.2 [196].

3.3.1 Generalised Bifrequency Transfer Function

Let $\mathbf{X}(\lambda)$ be the representation of the signal $\mathbf{x}(t)$ on the domain λ. If the input signal, $\mathbf{x}(t)$, is filtered by a linear operator, $\mathscr{L}$, then the output is given by $\mathbf{y}(t) =$

$\mathscr{L}\,\mathbf{x}(t)$, and represented on the domain λ using equations (3.18) as:

$$\mathbf{Y}(\lambda) = \int_{\mathsf{T}} \mathbf{y}(t)\, \mathsf{K}(t,\lambda)\, dt \qquad \text{and} \qquad \mathbf{y}(t) = \int_{\Lambda} \mathbf{Y}(\lambda)\, \mathsf{k}(\lambda,t)\, d\lambda \tag{3.37}$$

where $\mathbf{x}(t)$ is given by equation (3.18a). It is assumed that the initial state of the filters discussed throughout this chapter, is rest (zero initial conditions).

Theorem 8 (Transfer Function of a LTV System) *The impulse response of the system $\mathscr{L}$ may be written in the separable form:*

$$\mathsf{h}(t,\tau) = \iint_{\boldsymbol{\Lambda}} \mathsf{H}(\lambda,\hat{\lambda})\, \mathsf{k}(\lambda,t)\, \mathsf{K}(\tau,\hat{\lambda})\, d\lambda\, d\hat{\lambda}, \quad \forall (t,\tau) \in \mathbf{T} \tag{3.38a}$$

where $\{\mathsf{k}(\lambda,t),\ \mathsf{K}(t,\lambda),\ (t,\lambda) \in \mathsf{T}\times\Lambda\}$ *is a transform kernel pair given by* (3.10)*, and* $\mathsf{H}(\lambda,\hat{\lambda})$ *is the generalised bifrequency transfer function of $\mathscr{L}$, given by,*

$$\mathsf{H}(\lambda,\hat{\lambda}) = \iint_{\mathbf{T}} \mathsf{h}(t,\tau)\, \mathsf{K}(t,\lambda)\, \mathsf{k}(\hat{\lambda},\tau)\, dt\, d\tau, \quad \forall (\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda} \tag{3.38b}$$

The separable form of $\mathsf{h}(t,\tau)$ in Theorem 8 for linear time-varying differential equations is well known, as discussed in, for example, [417]. D'Angelo [84] gives a proof of this result avoiding the use of transform methods, and an outline of this proof is given in Appendix B.2.

PROOF. The proof of this theorem centres around the proof of equation (3.38a): given the definition of the GBTF, $\mathsf{H}(\lambda,\hat{\lambda})$, equation (3.38b) follows naturally from the definition of the integral transforms (3.8) as shown below.

A linear system $\mathscr{L}$ maps an input linear space $\mathscr{X}$ to an output linear space $\mathscr{Y}$ with the property $\mathscr{L}(a\mathbf{x}+b\mathbf{y}) = a\mathscr{L}\,\mathbf{x} + b\mathscr{L}\,\mathbf{y}$, where $\mathbf{x},\mathbf{y} \in \mathscr{X}$ and $a, b \in \mathbb{K}$ (the real/complex field); from (3.18):

$$\mathbf{y}(t) = \mathscr{L}\,\mathbf{x}(t) = \int_{\Lambda} \mathbf{X}(\lambda)\, \mathscr{L}\{\mathsf{k}(\lambda,t)\}\, d\lambda \tag{3.39}$$

where $\mathscr{L}$ operates on functions of t only. The representation of $\mathbf{x}(t)$ on the domain λ is given by $\mathbf{X}(\lambda)$ from equation (3.18a) and, by substituting (3.18a) into (3.39),

$$\mathbf{y}(t) = \iint_{\mathsf{T}\Lambda} \mathbf{x}(\tau)\, \mathscr{L}\{\mathsf{k}(\lambda,t)\}\, \mathsf{K}(\tau,\lambda)\, d\lambda\, d\tau \tag{3.40}$$

Comparing with the superposition integral,

$$\mathbf{y}(t) = \int_T h(t,\tau)\,\mathbf{x}(\tau)\,d\tau \tag{3.35}$$

the impulse response may be written as:

$$h(t,\tau) = \int_\Lambda \mathscr{L}\{k(\lambda,t)\}\,K(\tau,\lambda)\,d\lambda \tag{3.41}$$

Considering $\mathscr{L}\{k(\lambda,t)\}$ as a function of t, and treating λ as a parameter, it may be represented by the general transform (3.8b) with spectrum $H(\lambda,\hat{\lambda})$ defined by

$$\mathscr{L}\left\{k(\hat{\lambda},t)\right\} \triangleq \int_\Lambda H(\lambda,\hat{\lambda})\,k(\lambda,t)\,d\lambda \tag{3.42a}$$

Substituting (3.42a) into (3.41) gives (3.38a). Moreover, $\mathscr{L}\{k(\hat{\lambda},t)\}$ may also be expressed as:

$$\mathscr{L}\left\{k(\hat{\lambda},t)\right\} = \int_T h(t,\tau)\,k(\hat{\lambda},\tau)\,d\tau \tag{3.42b}$$

This expression is obtained by either inverting transform (3.41), with τ considered as a parameter, or by using the superposition integral, (3.35), with $\hat{k}(\lambda,t)$ as the input. By inverting (3.42a), with $\hat{\lambda}$ as a parameter, gives,

$$H(\lambda,\hat{\lambda}) = \int_T \mathscr{L}\left\{k(\hat{\lambda},t)\right\}\,K(t,\lambda)\,dt \tag{3.42c}$$

which, along with (3.42b), gives equation (3.38b). Note that the various methods of obtaining (3.42a) demonstrate the self-consistency of the definition of the GBTF. □

Theorem 9 (Spectral Convolution) *The relationship between the spectral functions* $\mathbf{Y}(\lambda)$ *and* $\mathbf{X}(\lambda)$ *is the convolution:*

$$\mathbf{Y}(\lambda) = \int_\Lambda H(\lambda,\hat{\lambda})\,\mathbf{X}(\hat{\lambda})\,d\hat{\lambda} \tag{3.43}$$

Gersho [114] provides equivalent forms of the spectral convolution in Theorem 9 for the case when the definition of the impulse response in equation (3.36) is used; a discussion of analogous results may also be found in recent work by Margrave [239].

PROOF. This follows by substituting (3.39) into the expression for $\mathbf{Y}(\lambda)$ in (3.37),

$$\mathbf{Y}(\lambda) = \iint_{T\Lambda} \mathbf{X}(\hat{\lambda})\, \mathscr{L}\left\{k(\hat{\lambda}, t)\right\}\, K(t, \lambda)\, d\hat{\lambda}\, dt \tag{3.44}$$

rearranging the order of integration, and using (3.42c). □

Now, consider the relationships derived above: in particular compare (3.42c) with (3.41), and (3.35) with (3.43). It can be seen that each of these pair of equations are equivalent in form, and differ only by the domain on which they operate: this emphasises that the transform is merely a change in basis. A special case emphasising that a change in basis can greatly simplify the filtering problem is for LTI systems.

Example 1 (Transfer Relations of LTI Systems). Consider a LTI filter; from (3.41):

$$h(t, \tau) = h(t - t_0, \tau - t_0) = \int_{\Lambda} \hat{k}(\lambda, t - t_0) K(\tau - t_0, \lambda)\, d\lambda\,, \quad \forall\, t_0 \in \mathbb{R} \tag{3.45}$$

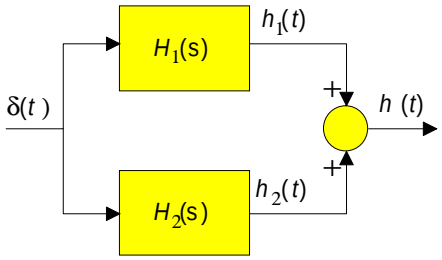
Setting $t_0 = \tau$, $h(t, \tau) = h(t - \tau) \triangleq h(t - \tau, 0)$, comparing with (3.41), and using a similar argument to that in section §3.2.4.3, it follows that the only other kernel satisfying this condition, with a unique reciprocal basis, is $k(\lambda, t) = \frac{1}{2\pi} e^{j\lambda t}$. Hence:

$$\begin{aligned} h(t, \tau) \equiv h(t - \tau) &= \frac{1}{2\pi} \int_{\Lambda} \mathscr{L}\left\{e^{j\lambda t}\right\}\, e^{j\lambda \tau}\, d\lambda \\ &= \frac{1}{2\pi} \int_{\Lambda} F(\lambda) e^{j\lambda(t-\tau)}\, d\lambda \end{aligned} \tag{3.46}$$

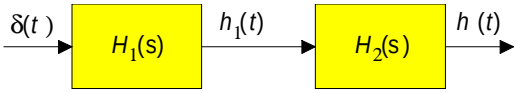
since $\mathscr{L}\left\{e^{j\lambda t}\right\} = F(\lambda)\, e^{j\lambda t}$. Hence, a LTI filter can only be expressed in the time or Fourier domain. Note that (3.38b) reduces to $H(\lambda, \hat{\lambda}) = F(\lambda)\, \delta(\lambda - \hat{\lambda})$, and (3.43) reduces to:

$$\mathbf{Y}(\lambda) = F(\lambda)\, \mathbf{X}(\lambda), \tag{3.47}$$

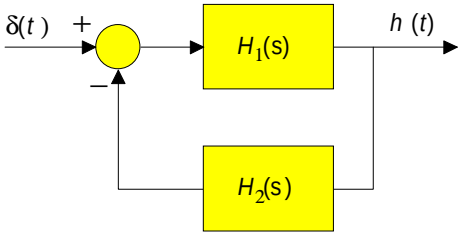
the familiar expression from LTI system theory. It is important to note that the converse of this theorem is, in general, false: if the input and output of a filter admit representation by the Fourier transform, the filter is not necessarily LTI. ⋈



(a) Parallel connection of two LTI systems



(b) Series connection of two LTI systems



(c) Feedback connection of two LTI systems

Figure 3.1: *Basic connections of two LTI systems having Laplace transfer functions* $H_1(s)$ *and* $H_2(s)$.

3.3.2 Block-Diagram Algebra

For LTI systems, the basic combinatorial relations of two LTI systems, having Laplace transfer functions $H_1(s)$ and $H_2(s)$, are shown in Figure 3.1, and are very useful in obtaining the transfer functions of a complex system. The impulse response and transfer function of each system are given by:

Parallel Combination:

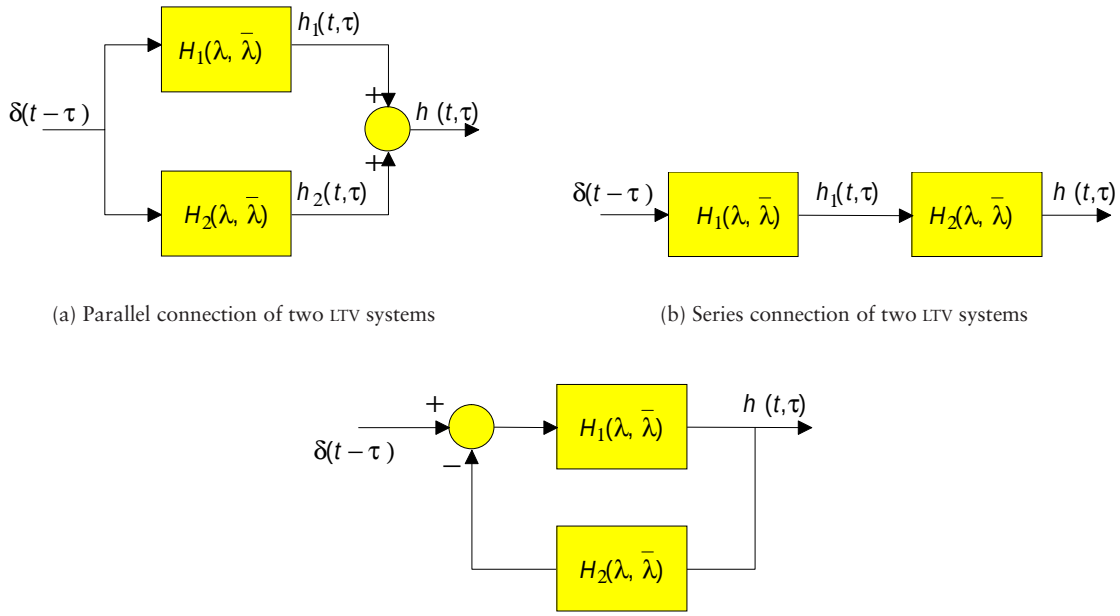
$$\begin{aligned} h(t) &= h_1(t) + h_2(t) \\ H(s) &= H_1(s) + H_2(s) \end{aligned} \tag{3.48a}$$

Series Combination:

$$\begin{aligned} h(t) &= h_1(t) \star h_2(t) \\ H(s) &= H_1(s)\, H_2(s) \end{aligned} \tag{3.48b}$$

Feedback Combination:

$$\begin{aligned} h(t) &= h_1(t) - [h_1(t) \star h_2(t)] \star h(t) \\ H(s) &= \frac{H_1(s)}{1 + H_1(s)\, H_2(s)} \end{aligned} \tag{3.48c}$$



(a) Parallel connection of two LTV systems

(b) Series connection of two LTV systems

(c) Feedback connection of two LTV systems

Figure 3.2: *Basic connections of two LTV systems having generalised bifrequency transfer functions* $H_1(\lambda, \hat{\lambda})$ *and* $H_2(\lambda, \hat{\lambda})$.

The set of corresponding relationships for combining LTV blocks is derived next to show that the GBTF is a natural extension of LTI transform theory. In each case, the LTV system, with impulse response $h(t, \tau)$, is composed of two blocks with impulse responses $h_1(t, \tau)$ and $h_2(t, \tau)$, and GBTFs $H_1(\lambda, \hat{\lambda})$ and $H_2(\lambda, \hat{\lambda})$, respectively.

3.3.2.1 Parallel Connection

It follows from Figure 3.2(a) and linearity that:

$$h(t, \tau) = h_1(t, \tau) + h_2(t, \tau) \tag{3.49a}$$

and, therefore, taking transforms using equation (3.38b) gives:

$$H(\lambda, \hat{\lambda}) = H_1(\lambda, \hat{\lambda}) + H_2(\lambda, \hat{\lambda}) \tag{3.49b}$$

3.3.2.2 Series Connection

Analysis of the cascade series connection of two linear time-varying systems shown in Figure 3.2(b), using the superposition integral (3.35), reveals that

$$h(t,\tau) = \int_T h_2(t,\hat{\tau})\, h_1(\hat{\tau},\tau)\, d\hat{\tau} \tag{3.50a}$$

Substituting $h_2(t,\hat{\tau})$ using equation (3.38a) and rearranging gives

$$h(t,\tau) = \iint_{\Lambda} H_2(\lambda_1,\lambda_2) \int_T h_1(\hat{\tau},\tau)\, k(\lambda_1,t)\, K(\hat{\tau},\lambda_2)\, d\hat{\tau}\, d\lambda_1\, d\lambda_2 \tag{3.50b}$$

Taking transforms of both sides of this equation using (3.38b), and simplifying using (3.10), gives:

$$H(\lambda,\hat{\lambda}) = \int_{\Lambda} H_2(\lambda,\bar{\lambda})\, H_1(\bar{\lambda},\hat{\lambda})\, d\bar{\lambda} \equiv H_2(\lambda,\hat{\lambda}) \star H_1(\lambda,\hat{\lambda}) \tag{3.50c}$$

where the convolution in (3.50c) is expected since, as noted in the previous section, the impulse response and the transfer function are related by a change in basis. It is important to notice, as discussed in D'Angelo [84] for Zadeh's transfer function, that the cascade series connection of two linear time-varying systems with impulse responses $h_1(t,\tau)$ and $h_2(t,\tau)$ is noncommutative: *i.e.* $H_2(\lambda,\hat{\lambda}) \star H_1(\lambda,\hat{\lambda}) \neq H_1(\lambda,\hat{\lambda}) \star H_2(\lambda,\hat{\lambda})$. The inverse generalised bifrequency transfer function (IGBTF) may be introduced by defining $H_1(\lambda,\hat{\lambda})$ and $H_2(\lambda,\hat{\lambda})$ to be inverses if, and only if, (iff):

$$H_1(\lambda,\hat{\lambda}) \star H_2(\lambda,\hat{\lambda}) = H_2(\lambda,\hat{\lambda}) \star H_1(\lambda,\hat{\lambda}) = 1, \quad \forall(\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda} \tag{3.51a}$$

The inverse of $H(\lambda,\hat{\lambda})$ may be written as $H^{-1}(\lambda,\hat{\lambda})$ and, as such, (3.51a) becomes:

$$H(\lambda,\hat{\lambda}) \star H^{-1}(\lambda,\hat{\lambda}) = H^{-1}(\lambda,\hat{\lambda}) \star H(\lambda,\hat{\lambda}) = 1, \quad \forall(\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda} \tag{3.51b}$$

3.3.2.3 Feedback Connections

Analysis of the cascade series connection of two LTV systems shown in Figure 3.2(c), using the superposition integral (3.35), shows that $h(t,\tau)$ can written in

the form:

$$h(t,\tau) = h_1(t,\tau) - \int_T h_1(t,\hat{\tau})\,\bar{h}(\hat{\tau},\tau)\,d\hat{\tau} \tag{3.52a}$$

where $$\bar{h}(t,\tau) = \int_T h_2(t,\bar{\tau})\,h(\bar{\tau},\tau)\,d\bar{\tau}$$

or, equivalently, in the form:

$$h(t,\tau) = h_1(t,\tau) - \int_T \hat{h}(t,\hat{\tau})\,h(\hat{\tau},\tau)\,d\hat{\tau} \tag{3.52b}$$

where $$\hat{h}(t,\tau) = \int_T h_1(t,\hat{\tau})\,h_2(\hat{\tau},\tau)\,d\hat{\tau}$$

Notice in (3.52b) that the integral equation in terms of $h(t,\tau)$, where $h_1(t,\tau)$ and $\hat{h}(t,\tau)$ are known, is a Volterra integral of the second kind, where either one of t or τ is considered as a parameter, and that its solution can be found by the method of successive approximation [291]. Using the results from equation (3.50c) in section §3.3.2.2, equation (3.52b) becomes:

$$H(\lambda,\hat{\lambda}) = H_1(\lambda,\hat{\lambda}) - \hat{H}(\lambda,\hat{\lambda}) \star H(\lambda,\hat{\lambda}) \tag{3.52c}$$

where $$\hat{H}(\lambda,\hat{\lambda}) = H_1(\lambda,\hat{\lambda}) \star H_2(\lambda,\hat{\lambda})$$

and where $\star$ is defined in (3.50c). The relationship introduced in (3.51b) could be used to obtain a form reminiscent of (3.48c). However, since the GBTF is simply a change of basis, in general, there is no simplification of the double convolution in (3.52b).

3.3.3 General System Function

As noted at the beginning of section §3.3, many authors of LTV system theory define the transfer function, or system function, in a much different way, *e.g.* see D'Angelo [84], Zadeh [409, 411, 417], Gersho [114] and Margrave [239]. Here, the system function depends on time, and is defined as:

$$G(\lambda,t) = \left.\frac{y(t)}{x(t)}\right|_{x(t)=k(\lambda,t)} \tag{3.53a}$$

where $y(t) = G(\lambda, t)\, k(\lambda, t)$ is the response of a zero-state system for the case when $k(\lambda, t)$ is taken as the input with the variable λ treated as a parameter. In such an instance, the input–output relation becomes

$$\mathbf{Y}(\lambda) = G(\lambda, t)\, \mathbf{X}(\lambda) \tag{3.53b}$$

where t is treated as a parameter, and $\mathbf{X}(\lambda)$ and $\mathbf{Y}(\lambda)$ are defined using the transform in (3.18). It may be shown [84] that $G(\lambda, t)$ is given by

$$G(\lambda, t) = \frac{1}{k(\lambda, t)} \int_T h(t, \tau)\, k(\lambda, \tau)\, d\tau \tag{3.53c}$$

Direct relationships between $G(\lambda, t)$ and the GBTF are:

$$H(\hat{\lambda}, \lambda) = \int_T G(\lambda, t)\, k(\lambda, t)\, K(t, \hat{\lambda})\, dt \tag{3.53d}$$

$$G(\lambda, t) = \frac{1}{k(\lambda, t)} \int_\Lambda H(\hat{\lambda}, \lambda)\, k(\hat{\lambda}, t)\, d\hat{\lambda} \tag{3.53e}$$

PROOF. Multiply both sides of (3.53c) by $k(\lambda, t)\, K(t, \hat{\lambda})$, and integral w. r. t. t over T. Then, comparing the right hand side of the resulting equation with (3.38b) gives the desired result. Similarly, (3.53e) follows using equations (3.42a) and (3.42b). □

This relationship is well known when $k(\lambda, t) = e^{\lambda t}$, and is often referred to as Zadeh's transfer function [409, 411]. The result is also known in the infinite duration discrete-time (continuous-spectral) case [227].[11] Although the definition of (3.53a) is useful in many applications, the GBTF tends to lead more naturally to a solution for the separability problem. Nevertheless, since there is a unique relationship between $G(\lambda, t)$ and $H(\lambda, \hat{\lambda})$, similar separability results could be obtained independently of the definition of the transfer function, provided such a unique relationship exists.

3.3.4 Randomly-Varying Systems

In the early 1950's, Zadeh [409, 412, 416] (see also [336]) proved some interesting results for a system whose impulse response, $h(t, \tau)$, is itself a stochastic process.

[11] Essentially this is Zadeh's transform with the Fourier kernel replaced by the $\mathscr{Z}$– transform kernel.

In the special case when the general system function, $\mathbf{G}(\omega, t)$, is a WSS process in time, for a given Fourier frequency ω, such that its correlation function is given by,

$$R_{GG}(\omega;\tau) = E\left[\mathbf{G}(\omega, t)\,\mathbf{G}^{*}(\omega, t+\tau)\right], \quad \forall t \in \mathbb{R} \tag{3.54}$$

Zadeh shows that if the input process, $\mathbf{x}(t)$, is WSS with power spectrum $\mathcal{P}_{xx}(\omega)$ *and* is independent of the system function, the output, $\mathbf{y}(t)$, is also WSS with power spectrum $\mathcal{P}_{yy}(\omega)$, where:

$$\mathcal{P}_{yy}(\omega) = \int_{\Omega} \mathcal{P}_{GG}(\omega, \hat{\omega})\,\mathcal{P}_{xx}(\hat{\omega})\,d\hat{\omega} \tag{3.55}$$

and $\mathcal{P}_{GG}(\omega, \hat{\omega})$ is the Fourier transform of $R_{GG}(\omega;\tau)\ e^{j\omega\tau}$ w. r. t. τ, *i.e.*

$$\mathcal{P}_{GG}(\omega, \hat{\omega}) = \int_{T} R_{GG}(\omega;\tau)\ e^{-j(\hat{\omega}-\omega)\tau}\,d\tau \tag{3.56}$$

Bello [31] extends Zadeh's work and gives a comprehensive discussion of stochastic linear systems. Similar results can be extended for the case when a particular realisation of the system function is characterised by the generalised bifrequency transfer function:

Theorem 10 (Input-Output Relations of Random Systems) *Consider a stochastic system with impulse response* $\mathbf{h}(t, \tau)$*, whose input is a general nonstationary random process,* $\mathbf{x}(t)$*, independent of* $\mathbf{h}(t, \tau)$*, and output given by* $\mathbf{y}(t)$*. If a particular realisation of the impulse response of the system given by* $h(t, \tau)$*, where* τ *is considered as a parameter, is asymptotically stable, such that*

$$\iint_{\mathbf{T}} |h(t, \tau)|^{2}\ dt\,d\tau < \infty \tag{3.57a}$$

then the relationship between the input and output autocorrelation function is:

$$R_{yy}(t, \tau) = \iint_{\mathbf{T}} R_{hh}\left(t, \hat{t}; \tau, \hat{\tau}\right) R_{xx}\left(\hat{t}, \hat{\tau}\right)\ d\hat{t}\,d\hat{\tau} \tag{3.57b}$$

where the autocorrelation function of the 2-D impulse response, $\mathbf{h}(t, \tau)$*, is given by:*

$$R_{hh}\left(t, \tau; \hat{t}, \hat{\tau}\right) \triangleq E\left[\mathbf{h}(t, \tau)\,\mathbf{h}^{*}(\hat{t}, \hat{\tau})\right] \tag{3.57c}$$

Moreover, the relationship between the input and output generalised power spectra

is given by:

$$\mathcal{P}_{yy}(\lambda,\hat{\lambda}) = \iint_{\Lambda} \mathcal{P}_{hh}(\lambda,\lambda';\hat{\lambda},\hat{\lambda}')\,\mathcal{P}_{xx}(\lambda',\hat{\lambda}')\,d\lambda'\,d\hat{\lambda}' \tag{3.57d}$$

where $$\mathcal{P}_{hh}(\lambda,\hat{\lambda};\lambda',\hat{\lambda}') = E\left[\mathbf{H}(\lambda,\hat{\lambda})\,\mathbf{H}^*(\lambda',\hat{\lambda}')\right] \tag{3.57e}$$

and $\mathbf{h}(t,\tau) \rightleftharpoons \mathbf{H}(\lambda,\hat{\lambda})$, *as defined by equation* (3.38b). *The similarity in the form of equations* (3.57b) *and* (3.57d) *again arises due to a simple change of basis.*

PROOF. The output of the stochastic system is given by the superposition integral

$$\mathbf{y}(t) = \int_{T} \mathbf{h}(t,\tau)\,\mathbf{x}(\tau)\,d\tau, \quad \forall t \in T \tag{3.35}$$

The autocorrelation function of the output may, hence, be written as:

$$E\left[\mathbf{y}(t)\,\mathbf{y}^*(\tau)\right] = \iint_{\mathbf{T}} E\left[\mathbf{h}(t,\hat{t})\,\mathbf{h}^*(\tau,\hat{\tau})\,\mathbf{x}(\hat{t})\,\mathbf{x}^*(\hat{\tau})\right]\,d\hat{t}\,d\hat{\tau} \tag{3.58a}$$

which, using the independence of $\mathbf{x}(t)$ and $\mathbf{h}(t,\tau)$, gives equation (3.57b) with the definition of the autocorrelation of the impulse response function in equation (3.57c).

Next, decomposing the impulse response using a 2-D stochastic transform analogous to equation (3.38a) yields:

$$\mathbf{h}(t,\tau) = \iint_{\Lambda} \mathbf{H}(\lambda,\hat{\lambda})\,k(\lambda,t)\,K(\tau,\hat{\lambda})\,d\lambda\,d\hat{\lambda}, \quad \forall (t,\tau) \in \mathbf{T} \tag{3.38a}$$

Substituting into equation (3.58a) gives:

$$\begin{aligned} R_{yy}(t,\tau) = \iiint_{\mathbf{T}\,\Lambda\,\Lambda} & E\left[\mathbf{H}(\lambda_1,\hat{\lambda}_1)\,\mathbf{H}^*(\lambda_2,\hat{\lambda}_2)\right] R_{xx}\left(\hat{t},\hat{\tau}\right) \\ & \times k(\lambda_1,t)\,K(\hat{t},\hat{\lambda}_1)\,k^*(\lambda_2,\tau)\,K^*(\hat{\tau},\hat{\lambda}_2)\,d\lambda_1\,d\hat{\lambda}_1\,d\lambda_2\,d\hat{\lambda}_2\,d\hat{t}\,d\hat{\tau} \end{aligned} \tag{3.58b}$$

Using equation (3.24a), by taking the generalised power spectrum of each side of equation (3.58b), rearranging the order of integration, and repeated application of the relationships in equations (3.10), gives the desired form in equation (3.57d) with the definition in (3.57e). □

3.3.5 Compatible Transforms

In Example 1, the Fourier transform arose as the natural choice for the spectral representation of a LTI system and, as such, equation (3.43) reduces to:

$$\mathbf{Y}(\lambda) = H(\lambda)\,\mathbf{X}(\lambda), \quad \lambda \in \Omega \subset \mathbb{R} \tag{3.47}$$

with convolution in the time domain becoming multiplication in the spectral domain. If, in general, a transform kernel, $k(t,\lambda)$, and an inverse can be found for a LTV system, such that (3.43) reduces to (3.47), then $k(t,\lambda)$ is the *eigenfunction* of the system [84]. It is also termed the *characteristic input*, and the integral transform (3.8) and λ-domain are termed, respectively, the *compatible transform* and *compatible domain* with respect to the particular system of interest.

3.3.5.1 Block Diagram Algebra

The block-diagram algebra for compatible domains are identical to those in equations (3.48) and can easily be proved by setting $H_i(\lambda,\hat{\lambda}) \equiv H_i(\lambda)\,\delta(\lambda-\hat{\lambda})$ in equations (3.49b), (3.50c) and (3.52b).

3.3.5.2 Compatible Transforms in Analysis

The Laplace or Fourier transforms are particularly useful in the analysis and synthesis of linear time-invariant systems. The advantages and values stemming from the use of these transforms hinge on the fact that the transformation of a linear differential equation with constant coefficients,

$$\mathscr{L}\,x(t) = \sum_{n=0}^{N} \alpha_n \frac{d^n y(t)}{dt^n} = u(t), \quad t \in \mathbb{R}^+ \tag{3.59}$$

yields an algebraic equation in the transform variables: *i.e.*

$$A(s)\,Y(s) = U(s) \tag{3.60}$$

where $u(t) \rightleftharpoons U(s)$, $y(t) \rightleftharpoons Y(s)$, $A(s) = \sum_{n=0}^{N} \alpha_n s^n$, with $s \in \mathbb{C}$. For linear time-varying systems, however, Laplace or Fourier transformations of the characterising linear differential equation with time-varying coefficients,

$$\mathscr{L}\, x(t) = \sum_{n=0}^{N} \alpha_n\, a_n(t) \frac{d^n y(t)}{dt^n} = u(t) \tag{3.61}$$

do not yield algebraic equations but, generally, another differential equation in the transform variables. Thus, in the analysis and synthesis of LTV systems, there is usually no advantage gained by using a Laplace or Fourier transformation, since solutions of difficult differential equations are still required. However, *compatible* transformations do yield an algebraic equation, simplifying the analysis of the system in the same way that taking Laplace transforms simplifies the analysis of LTI systems.

3.3.5.3 Compatible Transforms in Signal Separation

The simplifications in signal analysis that result from taking a compatible transform has two benefits in signal separation. First, suppose that a stochastic process, $\mathbf{d}(t)$, is known to be generated by a system described by the LTV differential equation,

$$\sum_{n=0}^{N} \alpha_n\, a_n(t)\, \frac{d^n}{dt^n}\, \mathbf{d}(t) = \sum_{m=0}^{M} \beta_m\, b_m(t)\, \frac{d^m}{dt^m}\, \mathbf{e}(t) \tag{3.62}$$

where the functions $\{a_n(t),\ (t,n) \in T \times \mathscr{N}\}$, $\mathscr{N} = \{0, \ldots, N\}$ and $\{b_m(t),\ (t,m) \in T \times \mathscr{M}\}$, $\mathscr{M} = \{0, \ldots, M\}$ are known, but the coefficients $\{\alpha_n,\ \beta_m,\ n \in \mathscr{N}, m \in \mathscr{M}\}$ and excitation, $\mathbf{e}(t)$, are unknown. Taking a compatible transform of the system equation yields:

$$\mathbf{D}(\lambda) = \frac{B(\lambda)}{A(\lambda)}\, \mathbf{E}(\lambda) \tag{3.63}$$

where $\mathbf{d}(t) \rightleftharpoons \mathbf{D}(\lambda)$, $\mathbf{e}(t) \rightleftharpoons \mathbf{E}(\lambda)$, $A(\lambda) = \sum_{n=0}^{N} \alpha_n \lambda^n$, and $B(\lambda) = \sum_{m=0}^{M} \beta_m \lambda^m$. Given two processes that are known to satisfy the form of (3.62), the processes are separable provided their compatible transform representations are disjoint in the compatible domain. Although the compatible transform domain must be determined from the structure of the signal, since the separation and separability problems are inherently under-constrained, it is acceptable to require knowledge

that the signal structures satisfy equation (3.62).

Second, suppose the sum $\mathbf{x}(t) = \mathbf{d}(t) + \mathbf{n}(t)$ is observed, and it is desired to estimate $\mathbf{d}(t)$ using the filter defined in equation (3.4). Taking a compatible transform w. r. t. the system defined by the impulse response, $h(t,\tau)$, of (3.4) yields the form of (3.47), rather than the more complicated form of (3.43). As such, if the representations of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ on this compatible λ-domain, given by $\mathbf{D}(\lambda)$ and $\mathbf{N}(\lambda)$ respectively, are disjoint, signal separability can be achieved with the polynomial $H(\lambda)$ representing a bandpass filter on the λ-domain. This aspect is investigated further in section §4.1.4. A disadvantage of this approach is that it only simplifies the separability problem for classes which are selected *a priori*; a method which uses the *structure* of the signal classes to find compatible classes is more powerful.

Example 2 (Generalised Differential Elements in Signal Separation). Suppose two signals $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are known to satisfy the system equation

$$\mathscr{L}\,\mathbf{d}(t) = \mathbf{e}_d(t) \quad \text{and} \quad \mathscr{L}\,\mathbf{n}(t) = \mathbf{e}_n(t) \tag{3.64}$$

where $\mathscr{L}$ is the first-order linear differential operator

$$\mathscr{L} = a_1(t)\,\frac{d}{dt} + a_0(t)$$

as shown in Figure 3.3 on page 72. It is shown in section §3.3.5.6 that the domain compatible with the linear operator $\mathscr{L}$ has transform kernels:

$$K(t,\lambda) = \phi(t)\,e^{-j\lambda\gamma(t)} \quad \text{and} \quad k(\lambda,t) = \frac{a_1(t)}{2\pi\phi(t)}\,e^{j\lambda\gamma(t)} \tag{3.65a}$$

where $\phi(t)$ and $\gamma(t)$ are given by:

$$\phi(t) = \frac{1}{a_1(t)}\exp\left\{\int_T \frac{a_0(t)}{a_1(t)}\,dt\right\} \quad \text{and} \quad \gamma(t) = \int_T \frac{1}{a_1(t)}\,dt \tag{3.65b}$$

If $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are disjoint in this domain, they are separable even though they overlap in the Fourier domain (see also [347]). ⋈

3.3.5.4 Classical Examples of Compatible Transforms

Details of taking compatible transforms may be found in, for example, the work of D'Angelo [84], Zadeh [414] and Aseltine [17]. There are a few well known

compatible transforms, *e.g.* the Laplace transform, compatible with systems characterised by LTI differential equations; the Mellin transform [127], compatible with the Euler-Cauchy equation; and the Hankel transform, compatible with the generalised Bessel equation. This section lists these classes of systems and their corresponding transform kernels. Solutions to general first and second-order LTV differential equations [17, 347] and a class of LTV discrete systems [194] are also presented.

The Laplace Transform The *characteristic input* for LTI systems is the well-known inverse transform kernel $e^{\lambda t}$ defining the Laplace transform. The transform equations are:

$$X(\lambda) = \int_{0^-}^{\infty} x(t)\, e^{-\lambda t}\, dt \tag{3.66a}$$

$$x(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(\lambda)\, e^{\lambda t}\, d\lambda \tag{3.66b}$$

where $c \in \mathbb{R}$. The Laplace transform is compatible with systems characterised by linear differential equations with constant coefficients, *i.e.* as typified by:

$$\mathscr{L}\, x(t) = \alpha_n \frac{d^n y(t)}{dt^n} + \alpha_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + \alpha_0 y(t) = u(t) \tag{3.59}$$

The Mellin Transform The inverse transform kernel corresponding to the Mellin transform is $t^{-\lambda}$. The transform equations are [127]:

$$X(\lambda) = \int_{0^-}^{\infty} x(t)\, t^{\lambda-1}\, dt \tag{3.67a}$$

$$x(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(\lambda)\, t^{-\lambda}\, d\lambda \tag{3.67b}$$

where $c \in \mathbb{R}$. The Mellin transform is compatible with systems characterised by the Euler-Cauchy equation:

$$\mathscr{L}\, x(t) = \alpha_n\, t^n \frac{d^n y(t)}{dt^n} + \alpha_{n-1}\, t^{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + \alpha_0\, y(t) = u(t) \tag{3.67c}$$

The Hankel Transform The inverse transform kernel of the Hankel transform is $\lambda J_n(\lambda t)$, where J_n denotes the Bessel function. The Hankel transform is:

$$X_n(\lambda) = \int_0^\infty x(t)\, t\, J_n(\lambda t)\, dt \tag{3.68a}$$

$$x(t) = \frac{1}{2\pi j} \int_0^\infty X_n(\lambda)\, \lambda\, J_n(\lambda t)\, d\lambda \tag{3.68b}$$

The Hankel transform is compatible with systems characterised by the generalised Bessel equation:

$$\mathscr{L}\, x(t) = \left[\frac{d^2}{dt^2} + \frac{1}{t}\frac{d}{dt} - \left(\frac{n^2}{t^2}\right) \pm a^2\right]^N x(t) = f(t) \tag{3.68c}$$

3.3.5.5 Transform Kernel of a Continuous-Time LTV System

An approach to obtaining the compatible transform for a large class of systems hinges on the assumption that a system can be decomposed into a combination of *fundamental building blocks* for which the compatible domain is known; the compatible transform for the entire system is the same as for the fundamental building block. This approach is discussed in section §3.3.5.6 for continuous LTV systems, and in section §3.3.5.7 for discrete LTV systems.

Theorem 11 (Compatible Transform Kernel) *The direct transform kernel,* $K(t, \lambda)$*, of the integral transform defined in equations* (3.8) *which transforms a given system equation, characterised by the linear operator* $\mathscr{L}$*, into an algebraic equation in one variable, such that*

$$y(t) = \mathscr{L}\, x(t) \rightleftharpoons Y(\lambda) = \int_T \mathscr{L}\, x(t)\, K(t, \lambda)\, dt = H(\lambda) X(\lambda) \tag{3.69a}$$

where $X(\lambda)$ *is the representation of* $x(t)$ *on the domain* λ*, is given by the solution of:*

$$\mathscr{L}^* K(t, \lambda) = H(\lambda)\, K(t, \lambda) \tag{3.69b}$$

for some function $H(\lambda)$ *in* λ. $\mathscr{L}^*$ *is the* adjoint operator *to* $\mathscr{L}$, *and* all *initial conditions are zero. The solution to equation* (3.69b) *is compatible with all systems governed by the differential equation:*

$$\left(1 + \sum_{p=1}^{P} \alpha_p \mathscr{L}^p\right) y(t) = \left(\sum_{q=0}^{Q} \beta_q \mathscr{L}^q\right) u(t) \tag{3.69c}$$

The formulation for the direct transform kernel in equation (3.69b) is desirable since, unlike in equation (3.69a), there is no dependence on the signals within the system, and is an equation in the transform kernel only.

PROOF. The first part of the proof relies on finding a solution to equation (3.69a) for some function of $H(\lambda)$:

$$H(\lambda)X(\lambda) = \int_T \mathscr{L}\, x(t)\, K(t,\lambda)\, dt \tag{3.69a}$$

Here, $X(\lambda)$ is the representation of $x(t)$ on the domain λ, given by equation (3.8a), and *all* initial conditions are assumed zero. Letting $\langle u, v\rangle$ denote the scalar product of the two functions u and v given by:

$$\langle u, v\rangle = \int_T u v^*\, dt \tag{3.70a}$$

then equations (3.8a) and (3.69a) may be written as:

$$X(\lambda) = \int_T x(t)\, K(t,\lambda)\, dt = \langle x(t), K(t,\lambda)\rangle \tag{3.8a}$$

$$H(\lambda)\, X(\lambda) = \int_T \mathscr{L}\, x(t)\, K(t,\lambda)\, dt = \langle \mathscr{L}x(t), K(t,\lambda)\rangle \tag{3.69a}$$

Now, if $\mathscr{L}$ is a linear operator, the *adjoint operator* to $\mathscr{L}$, designated by $\mathscr{L}^*$, exists and is uniquely defined by the expression [84, 291, 347]:

$$\langle \mathscr{L}u, v\rangle = \langle u, \mathscr{L}^* v\rangle \tag{3.70b}$$

If therefore follows from equation (3.69a) that the adjoint operator $\mathscr{L}^*$ satisfies:

$$H(\lambda)\, X(\lambda) = \langle \mathscr{L}x(t), K(t,\lambda)\rangle = \langle x(t), \mathscr{L}^* K(t,\lambda)\rangle = \int_T x(t)\, \mathscr{L}^*\, K(t,\lambda)\, dt \tag{3.70c}$$

Comparing this expression with (3.8a) gives the requirement in equation (3.69b).

The second part of the proof, showing the solution to equation (3.69b) is compatible with all systems governed by the differential equation in (3.69c), follows by noting that since $\mathscr{L}\, x(t) \rightleftharpoons H(\lambda)\, X(\lambda)$:

$$\begin{aligned}\mathscr{L}\,\{\mathscr{L}\, x(t)\} &\rightleftharpoons \langle \mathscr{L}\,\{\mathscr{L}\, x(t)\}, K(t,\lambda)\rangle \equiv \langle \mathscr{L}\, x(t), \mathscr{L}^*\, K(t,\lambda)\rangle \\ &\rightleftharpoons \langle \mathscr{L}\, x(t), H(\lambda)\, K(t,\lambda)\rangle = H(\lambda)\langle \mathscr{L}\, x(t), K(t,\lambda)\rangle = [H(\lambda)]^2\, X(\lambda)\end{aligned}$$

where (3.69b) has been used. It follows that this result can be generalised to give:

$$\mathscr{L}^n\, x(t) \rightleftharpoons [H(\lambda)]^n\, X(\lambda) \tag{3.70d}$$

By taking the transform of equation (3.69c) w. r. t. the kernel function, $K(t,\lambda)$, an algebraic equation in λ is obtained:

$$Y(\lambda) = \frac{\sum_{q=0}^{Q} \beta_q\, H^q(\lambda)}{1 + \sum_{p=1}^{P} \alpha_p\, H^p(\lambda)}\, U(\lambda) \tag{3.71}$$

where $y(t) \rightleftharpoons Y(\lambda)$ and $u(t) \rightleftharpoons U(\lambda)$. A convenient choice of $H(\lambda) = \lambda$, $\lambda \in \mathbb{C}$ or $H(\lambda) = \lambda^{-1}$, $\lambda \in \mathbb{C}$. □

3.3.5.6 A Class of Continuous LTV Systems

The following sections show how (3.69b) can be used to obtain the kernels functions for first and second-order differential equations with time-varying coefficients. Further approaches for obtaining kernels for systems with time-varying differential equations are discussed in [270]. Note that if the linear operator $\mathscr{L}$ is of the form

$$\mathscr{L} \equiv a_n(t)\, \frac{d^n}{dt^n} + \cdots + a_1(t)\, \frac{d}{dt} + a_0(t) \tag{3.72a}$$

then it can be shown that the adjoint system is characterised by the linear differential equation [84, pp. 308]:

$$\mathscr{L}^* \equiv (-1)^n \frac{d^n}{dt^n} \left[a_n(t) + \cdots - \frac{d}{dt}\, [a_1(t) + a_0(t) \right. \tag{3.72b}$$

where each term should be considered as an individual operator.

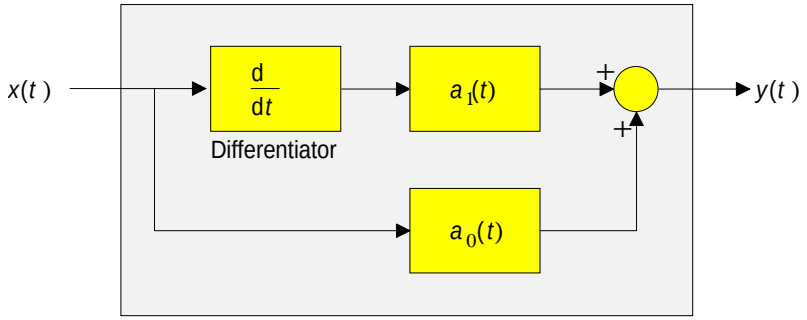


Figure 3.3: *Generalised differential element.*

First-Order Generalised Differentiator Consider the building block shown in Figure 3.3 which is governed by the first-order linear differential equation given by:

$$\mathscr{L}\, x(t) = a_1(t)\frac{dx(t)}{dt} + a_0(t)x(t) = f(t) \tag{3.73}$$

where $a_0(t)$ is non-zero and all the initial conditions are zero. This differential operator has been used in modelling time-varying electrical components [27, 29]. The adjoint operator $\mathscr{L}^*$ for the first-order linear operator $\mathscr{L}$ is given by equation (3.72b) and, therefore, equation (3.69b) becomes:

$$\mathscr{L}^*\, K(t,\lambda) = -\frac{d}{dt}\left[a_1(t)\, K(t,\lambda)\right] + a_0(t)\, K(t,\lambda) = H(\lambda)\, K(t,\lambda)$$

which may be written in the form:

$$\frac{1}{a_1(t)\, K(t,\lambda)}\,\frac{d}{dt}\left[a_1(t)\, K(t,\lambda)\right] = \frac{a_0(t) - H(\lambda)}{a_1(t)}$$

This is in integrable form and, therefore,

$$\ln\left[a_1(t)\, K(t,\lambda)\right] = \int_T \frac{a_0(t) - H(\lambda)}{a_1(t)}\, dt \quad \Rightarrow \quad K(t,\lambda) = \phi(t)\, e^{H(\lambda)\gamma(t)} \tag{3.74}$$

where $\phi(t)$ and $\gamma(t)$ are defined in equations (3.65) below. To develop the inverse kernel, it is noted that the kernels must be orthogonal, and hence must satisfy equations (3.10). Noting the orthogonality of exponential functions, it follows that there is a constraint on $H(\lambda)$ and should be chosen as $H(\lambda) = -j\lambda$.[12] Ergo, the

[12] Belal and Shenoi [29] note that if $a_0(t) = \dot{a}_1(t)$, $\phi(t) \equiv 1$, the operator $\mathscr{L}$ defined in (3.73) represents the voltage-current relationship of a time-varying inductor $a_1(t)$, with $\hat{\lambda} = j\lambda$ repre-

direct the inverse transform kernels are given by:

$$K(t,\lambda) = \phi(t)\, e^{-j\lambda\gamma(t)} \quad \text{and} \quad k(\lambda,t) = \frac{a_1(t)}{2\pi\phi(t)}\, e^{j\lambda\gamma(t)} \tag{3.65a}$$

where $\phi(t)$ and $\gamma(t)$ are given by:

$$\phi(t) = \frac{1}{a_1(t)} \exp\left\{\int_T \frac{a_0(t)}{a_1(t)}\, dt\right\} \quad \text{and} \quad \gamma(t) = \int_T \frac{1}{a_1(t)}\, dt \tag{3.65b}$$

This kernel is compatible with systems governed by the differential equation [27, 29]:

$$\left(1 + \sum_{p=1}^{P} \alpha_p \mathscr{L}^p\right) y(t) = \left(\sum_{q=0}^{Q} \beta_q \mathscr{L}^q\right) u(t) \tag{3.69c}$$

Unfortunately, elegant solutions are not possible for higher order systems, as shown in the next section.

Second-Order System Consider a system which is governed by the second-order linear differential equation given by:

$$a_2(t)\frac{d^2x(t)}{dt^2} + a_1(t)\frac{dx(t)}{dt} + a_0(t)x(t) = f(t) \tag{3.75}$$

where $a_0(t)$ is non-zero and all the initial conditions are zero. It can be shown [17, 84] that the transform kernel for this second-order system is given in the form:

$$K(t,\lambda) = g(t)\,\hat{K}(t,\lambda) \quad \text{where} \quad \begin{cases} g(t) = \exp\left\{\int \left[\dfrac{b_1(t) - \dot{b}_2(t)}{b_2(t)}\right] dt\right\} \\ b_2(t) = \dfrac{a_2(t)}{a_0(t)}, \qquad b_1(t) = \dfrac{a_1(t)}{a_0(t)} \end{cases} \tag{3.76}$$

where $\hat{K}(t,\lambda)$ satisfies the differential equation

$$b_2(t)\frac{d^2\hat{K}(t,\lambda)}{dt^2} + b_1(t)\frac{d\hat{K}(t,\lambda)}{dt} - H(\lambda)\,\hat{K}(t,\lambda) = 0 \tag{3.77}$$

---

senting the impedance. As a particular example, if $\hat{\lambda} \triangleq -j\lambda$, the direct kernel in (3.65a) reduces to $K(\hat{\lambda},t) = (1+t)^{-\hat{\lambda}}$ for systems characterised by the derivative operator $\mathscr{L} = (1+t)\frac{d}{dt} + 1 \equiv \frac{d}{dt}[(1+t)$ [29, 84].

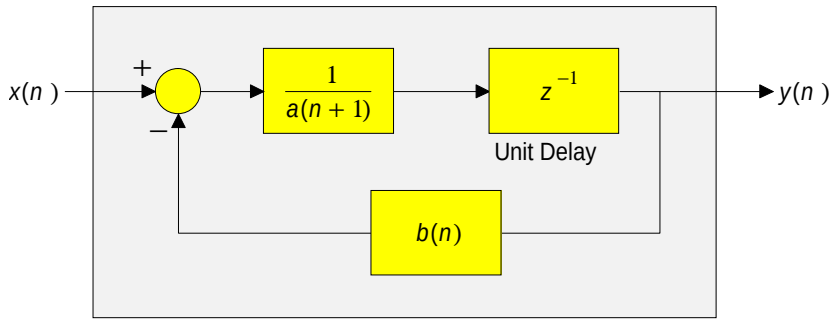


Figure 3.4: *Fundamental time-varying element.*

for some polynomial $H(\lambda)$. Therefore, knowledge of the solution of a linear differential equation is essential for determining the desired kernel of the integral transform under which transformation of the differential equation of interest yields an algebraic equation. Thus, although this method for determining the compatible transform kernel produced an elegant general result for the first-order system, it is clear that the compatible transform method is most advantageous in situations where the primary interest is in finding the responses of a single system, or particular type of system which is composed of a number of fixed building blocks, to a large variety of inputs otherwise, for each system or model of a stochastic process, solutions to equation (3.69b) will have to be obtained [17].

3.3.5.7 A Class of Discrete LTV Systems

Compatible transforms have also been developed for LTV discrete systems; for example, Klafter [194] has discussed the time-varying element shown in Figure 3.4 with input–output relationship:

$$y(n) = \frac{1}{a(n+1)\,z + b(n)}\,x(n), \quad n \in \mathcal{N} \tag{3.78}$$

where $a(n+1) \neq 0$, $\mathcal{N} = \{0, \ldots, \infty\}$ and $z^{-i}\,x(n) = x(n-i)$ is the delay operator. Cascade, parallel, fixed gain and feedback combinations of this basic LTV building block can, as shown in Figure 3.1, be used to model a wide range of discrete-time LTV systems. Klafter [194] has shown that the transforms compatible with systems

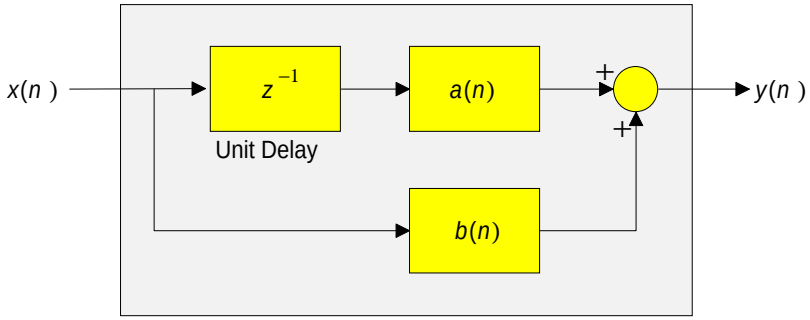


Figure 3.5: *Generalised delay element.*

composed of this building block are defined by the transform:[13]

$$X(\lambda) = \sum_{n=n_0}^{\infty} x(n)\, K(n,\lambda), \qquad K(n,\lambda) = \frac{\prod_{r=0}^{n_0} [\lambda - b(r)]}{\prod_{r=0}^{n} [\lambda - b(r)]} \prod_{r=0}^{n} a(r),\ \lambda \in \mathbb{C} \tag{3.79a}$$

$$x(n) = \frac{1}{2\pi j} \oint_{\mathscr{C}} X(\lambda)\, k(\lambda,n)\, d\lambda, \quad k(\lambda,n) = \frac{\prod_{r=0}^{n-1} [\lambda - b(r)]}{\prod_{r=0}^{n_0} [\lambda - b(r)]} \prod_{r=n_0}^{n} a(r),\ n \in \mathscr{N} \tag{3.79b}$$

where the contour, $\mathscr{C}$, encloses all the poles of

$$X(\lambda) \frac{\prod_{r=0}^{n-1} [\lambda - b(r)]}{\prod_{r=0}^{n_0} [\lambda - b(r)]} \tag{3.80}$$

and n_0 is an arbitrary starting time with $n \geq n_0$. Furthermore, the discrete analogue to equation (3.73) given as the first-order operator defined by

$$\mathscr{P}\, x(n) = a_1(n)\, x(n-1) + a_0(n)\, x(n) \tag{3.81a}$$

[13] Note that the fraction $\frac{\prod_{r=0}^{n_0} [\lambda - b(r)]}{\prod_{r=0}^{n} [\lambda - b(r)]}$ in (3.79b) is so written in order to deal with the case $n = n_0$. The corresponding fraction in (3.79a) is so written to show where computational savings can be made by reusing common terms.

has been investigated by Belal and Shenoi [28], and a transform compatible with the special case of time-varying difference equations

$$\left(1 + \sum_{p=1}^{P} \alpha_p \mathscr{P}^p\right) x(n) = \left(\sum_{q=0}^{Q} \beta_q \mathscr{P}^q\right) f(n) \tag{3.81b}$$

has kernel function

$$K(n,\lambda) = \frac{1}{a_1(n)} \prod_{i=0}^{n-1} \frac{H(\lambda) - a_0(i)}{a_1(i)} \tag{3.81c}$$

for some function $H(\lambda)$. Note the similarity of this expression with that for the kernel of the continuous first-order differential operator in equation (3.65a). Although it is difficult to obtain a general inverse kernel, Belal and Shenoi [28] discuss some special cases for the functions $a_1(n)$ and $a_0(n)$. Further uses of compatible transforms in discrete LTV systems have been discussed by Pei and Kiang [288].

3.4 Chapter Summary

The analysis of *linear time-varying* systems and *nonstationary stochastic processes* is less developed than for *linear time-invariant* systems and *stationary stochastic processes*, and this chapter has introduced some fundamental concepts which go some way towards narrowing this disparity. The problem of representing both deterministic and stochastic signals, initially expressed in the time domain, in another general domain has been considered. A general and consistent method for representing signals on an arbitrary domain using the method of basis transformations is discussed. In this method, the signals are represented on arbitrary domains by either integral transforms in continuous time, or discrete matrix transforms in discrete time. Using these transforms, the deterministic power spectrum of a stochastic process has been introduced. These results are used throughout Chapter 4 to develop a criterion for the separation of nonstationary signals with perfect precision.

Name	Compatible System Operator	Forward Transform kernel $K(t,\lambda)$	Inverse Transform kernel $k(\lambda,t)$	Defined in Equation:
Laplace	$\frac{d}{dt}$	$e^{-\lambda t}$	$\frac{1}{2\pi j}e^{\lambda t}$	(3.66) on page 68
Mellin [127]	$t\frac{d}{dt}$	$t^{\lambda-1}$	$\frac{1}{2\pi j}t^{-\lambda}$	(3.67) on page 68
Hankel	$\frac{d^2}{dt^2}+\frac{1}{t}\frac{d}{dt}-\left(\frac{n^2}{t^2}\right)\pm a^2$	$t\,J_n(\lambda t)$	$\frac{1}{2\pi j}\lambda\,J_n(\lambda t)$	(3.68) on page 69
Generalised Differentiator	$a_1(t)\frac{d}{dt}+a_0(t)$	$\frac{1}{a_1(t)}e^{\int_T \frac{a_0(t)-j\lambda}{a_1(t)}\,dt}$	$\frac{1}{2\pi}e^{\int_T \frac{-a_0(t)+j\lambda}{a_1(t)}\,dt}$	(3.65) on page 67
Generalised Delay	$a_1(n)\,z^{-1}+a_0(n)$	$\frac{1}{a_1(n)}\prod_{i=0}^{n-1}\frac{H(\lambda)-a_0(i)}{a_1(i)}$	—	(3.81) on page 75
Klafter [194]	$\frac{1}{a(n+1)\,z+b(n)}$	$\frac{\prod_{r=0}^{n_0}[\lambda-b(r)]}{\prod_{r=0}^{n}[\lambda-b(r)]}\prod_{r=0}^{n}a(r)$	$\frac{\prod_{r=0}^{n-1}[\lambda-b(r)]}{\prod_{r=0}^{n_0}[\lambda-b(r)]}\prod_{r=n_0}^{n}a(r)$	(3.79) on page 75

Table 3.1: *Summary of some common compatible transforms.*

4

Separation Techniques

TECHNIQUES were developed in the previous chapter for analysing nonstationary and linear time-varying systems. Using these results, section §4.1 investigates the form of a separability criterion by deriving the so-called *ideal filter*, whose output is the desired signal, $\mathbf{d}(t)$, when the input is the observed signal, $\mathbf{x}(t) = \mathbf{d}(t) + \mathbf{n}(t)$, composed of the sum of the desired signal and noise signal, $\mathbf{n}(t)$. A general solution for the nonstationary Wiener filter is outlined in section §4.2, and the subsequent question then discussed is: Which class of signals satisfy this separability criterion? In the rest of this chapter, the problem of determining signal classes that fit this criterion is considered. Some examples of signal classes which would 'classically' be considered inseparable are given in section §4.6.

4.1 The Ideal Filter

In many areas of Information Engineering, the concept of the bandpass filter is commonly understood to be a system which passes, without distortion, all Fourier

frequency components falling in a certain frequency range, the filter's pass-band, and rejects all frequency components falling outside this range, the stop-band; such filters can be realised asymptotically using LTI networks. The ideal bandpass filter is often referred to as an *ideal filter*. This concept is particular useful in the frequency division multiplexing (FDM) modulation scheme where signals are modulated to lie in a division of the frequency domain unused by other signals. Recovery of a specific signal is thus a matter of using an ideal filter to pass the desired region of the frequency band, and reject all other components. In time division multiplexing (TDM), a signal is recovered by designing a receiver which passes, without distortion, all signal components falling in a certain time window, and rejects all signal components falling outside this window. Such LTV filters can be realised as an ideal filter using a 'switch', exhibiting the same characteristics as the bandpass filter in the Fourier spectral domain, except the filter now operates in the temporal domain. These ideas suggest it should be possible to define an ideal filter on any arbitrary domain which passes, without distortion, all *generalised frequency* components falling in a certain range and rejects all others. As in section §3.3, the structure of the ideal filter is developed in terms of the transform kernels used to represent the signals which the filter operates upon.

4.1.1 Definition

The generalised ideal filter was first proposed by Zadeh in 1952 [413–415, 418]. In the analysis which follows, it is useful to take advantage of 'functional analysis' (see, for example, [291, Appendix 1], [408]), which allows the elements of a normed vector space (a *Banach* space) to represent functions. The advantage of this viewpoint is partly that an element of a Banach space is regarded as a vector, or point in space, which is conceptually simpler than thinking of it as a function. Only when interpreting the results of the functional analysis are vectors considered functions in their own right.

Consider two sets of signals $\{\mathbf{d}(t),\ t \in \mathsf{T}\}$ and $\{\mathbf{n}(t),\ t \in \mathsf{T}\}$. A particular realisation of the signal $\mathbf{d}(t)$, $d(t)$, may be represented as a vector $\mathbf{d}$ in a Banach space $\mathscr{D}$, and a particular realisation of $\mathbf{n}(t)$, $n(t)$, may also be represented as a vector $\mathbf{n}$ in a Banach space $\mathscr{N}$. Since the property of the ideal filter is that it will pass, without distortion, any signal that belongs to a particular class of signals, $\mathscr{D}$, and rejects all signals which do not belong to that class, and therefore lies in $\mathscr{N}$, the filter must satisfy the following definition:

Definition 6 (Ideal Filter). An operator $\mathscr{L}$ is said to be ideal if

$$\mathscr{L}(\mathbf{d}+\mathbf{n}) = \mathbf{d}, \quad \forall (\mathbf{d},\mathbf{n}) \in \mathscr{D} \times \mathscr{N} \tag{4.1}$$

Assuming both $\mathscr{D}$ and $\mathscr{N}$ contain the null set $\{\emptyset\}$, which represents a function of null value, then:

$$\mathscr{L}\,\mathbf{d} = \mathbf{d}, \quad \forall\, \mathbf{d} \in \mathscr{D} \tag{4.2a}$$

$$\mathscr{L}\,\mathbf{n} = \mathbf{0}, \quad \forall\, \mathbf{n} \in \mathscr{N} \tag{4.2b}$$

The manifolds $\mathscr{D}$ and $\mathscr{N}$ are often referred to as the *acceptance* and *rejection* manifolds of $\mathscr{L}$ respectively. It is noted that (4.2) does not necessarily imply (4.1) except in the case of linear filters. A nonlinear rectifier is an example of a filter satisfying equations (4.2), but is not an ideal filter [418]. ♢

Example 3 (Nonlinear Non-ideal Filters). A nonlinear rectifier is a system which passes only positive signals. If $\mathscr{D}$ is the set of all positive signals, and $\mathscr{N}$ is the set of all negative signals (assume both $\mathscr{D}$ and $\mathscr{N}$ contain the null value), equations (4.2) are satisfied. However, (4.1) does not hold since there exists $\mathbf{d}(\mathrm{t})$ and $\mathbf{n}(\mathrm{t})$ such that $\mathbf{d}(\mathrm{t}) + \mathbf{n}(\mathrm{t}) \in \mathscr{D}$. Hence, a nonlinear rectifier is not an ideal filter. ⋈

Theorem 12 (Properties of the Ideal Filter) *The following three properties result from the definition of an ideal filter.*

1. *Let $\mathscr{X}$ be the input space to the operator $\mathscr{L}$. If $\mathscr{L}$ is an ideal filter, it is idempotent in $\mathscr{X}$. Hence*

$$\mathscr{L}\{\mathscr{L}\,\mathbf{x}\} = \mathscr{L}\,\mathbf{x} \tag{4.3a}$$

2. *A linear idempotent filter is an ideal filter.*

3. *If $\{\mathbf{d}(\mathrm{t})\}$ and $\{\mathbf{n}(\mathrm{t})\}$ are two signal classes spanning the subspaces $\mathscr{D}$ and $\mathscr{N}$, they are separable if, and only if, $\mathscr{D}$ and $\mathscr{N}$ are virtually disjoint:*

$$\mathscr{D} \cap \mathscr{N} = \{\emptyset\} \tag{4.3b}$$

implying

$$\mathscr{D} \oplus \mathscr{N} \oplus \mathscr{V} = \mathscr{X} \tag{4.3c}$$

where $\mathscr{V}$ is the remaining subspace of $\mathscr{X}$, and $\oplus$ denotes the orthogonal direct sum.

PROOF. A proof is given below for completeness, and can also be found in [347, 418]; it uses the properties of the ideal filter as defined in equations (4.1) and (4.2):

1. Let $\mathscr{L}$ be an ideal filter for the input space $\mathscr{X} = \mathscr{D} \cap \mathscr{N} = \{\mathbf{d} + \mathbf{n},\ \forall \mathbf{d} \in \mathscr{D},\ \forall \mathbf{n} \in \mathscr{N}\}$ where $\mathscr{D}$ and $\mathscr{N}$ are its acceptance and rejection spaces respectively. Then, equations (4.1) and (4.2) give:

$$\mathscr{L}\{\mathscr{L}\,\mathbf{x}) = \mathscr{L}\{\mathscr{L}\,(\mathbf{d} + \mathbf{n})\} = \mathscr{L}\,\mathbf{d} = \mathbf{d} = \mathscr{L}\,\mathbf{x} \tag{4.4}$$

2. Let the linear operator $\mathscr{L}$ be idempotent in $\mathscr{X}$ and assume that $\mathbf{y} = \mathscr{L}\,\mathbf{x}$ for some $\mathbf{x} \in \mathscr{X}$. If $\mathscr{L}$ is an idempotent operator then, from equation (4.3a), $\mathscr{L}\{\mathscr{L}\,\mathbf{x}\} = \mathscr{L}\,\mathbf{x}$ and, therefore, $\mathscr{L}\,\mathbf{y} = \mathbf{y}$. Hence, since $\mathscr{L}$ is linear, $\mathscr{L}\,(\mathbf{x} - \mathbf{y}) = \mathscr{L}\,\mathbf{x} - \mathscr{L}\,\mathbf{y} = 0$, and $\mathscr{L}$ is an ideal filter which separates the signal classes:

$$\mathscr{D} = \{\mathbf{d} \ :\ \mathbf{d} = \mathscr{L}\,\mathbf{x}, \quad \mathbf{x} \in \mathscr{X}\}$$
$$\mathscr{N} = \{\mathbf{n} \ :\ \mathbf{n} = \mathbf{x} - \mathbf{d}, \mathbf{x} \in \mathscr{X}, \mathbf{d} \in \mathscr{D}\}$$

3. **Necessity**

   Let the signal classes $\mathscr{D} = \{\mathbf{d}(t)\ , t \in \mathsf{T}\}$ and $\mathscr{N} = \{\mathbf{n}(t)\ , t \in \mathsf{T}\}$ be separable; then there exists a filter which satisfies equations (4.2). These equations imply that the only common element in $\mathscr{D}$ and $\mathscr{N}$ is the null element in $\mathscr{X}$ and, therefore, the subspaces of $\mathscr{D}$ and $\mathscr{N}$ are virtually disjoint.

   Sufficiency

   Let the subspaces $\mathscr{D}$ and $\mathscr{N}$ be virtually disjoint; then there exists a linear operator which maps $\mathscr{D}$ to itself and $\mathscr{N}$ to zero. This operator, $\mathscr{L}$, is called a projector, and defines the projection of $\mathscr{X}$ on $\mathscr{D}$ along $\mathscr{N}$. Thus, $\mathscr{L}$ satisfies equations (4.2) and, by linear superposition, satisfies equation (4.1). As such, the signal classes $\mathscr{D} = \{\mathbf{d}(t),\ t \in \mathsf{T}\}$ and $\mathscr{N} = \{\mathbf{n}(t),\ t \in \mathsf{T}\}$ are separable. □

Note that the last part of the proof introduces the term 'projector'. The properties of projections in functional analysis can be applied directly to the study of ideal filters [418] and eases their analysis. In this thesis, however, an analytic formulation is developed, and projection theory is not required for the analysis.

Example 4 (Idempotent Filters). Note that part 1 of Theorem 12 means a tandem combination of two such filters is equivalent to $\mathscr{L}$. As an example, two cascaded

ideal low pass filters (or switches) are identical to a single low pass filter (or switch), assuming the acceptance regions are identical. Note, further, that if a nonlinear filter is idempotent, it is not necessarily an ideal filter. For example, the nonlinear rectifier of Example 3 is idempotent but, as previously noted, it is not an ideal filter.

⋈

4.1.2 Ideal Filtration of Random Signals

The definition of an ideal filter applies to random signals and, to clarify this, consider the following two interpretations of a stochastic process:

1. A stochastic processes, $\mathbf{x}(t)$, is a rule for assigning to every outcome, ζ, of an experiment, a function $\mathbf{x}(t, \zeta)$ [81,286]. Thus, a stochastic process is a family of time-domain functions depending on the parameter ζ or, equivalently, a function of t and ζ. The ζ-domain is the set of all experimental outcomes and the t-domain is the set $\mathbb{R}$ of real numbers. The space spanned by the set of functions $\mathbf{x}(t, \zeta)$ is called the ensemble space of the process. It follows that the results of section §4.1.1 apply to stochastic signals when the acceptance or rejection signal space of an ideal filter is chosen to equal the ensemble space.

2. As introduced in section §3.2.3, a stochastic process may be represented by a linear combination of deterministic functions, with random variables as the coefficients: the process therefore spans a subspace. The linear ideal filter can be used with stochastic processes if the acceptance and rejection manifolds are chosen as these subspaces.

Since an ideal filter can separate stochastic processes, a set of deterministic signals may be interpreted as a set of realisations of some stochastic process. Therefore, the separability of two deterministic signal classes becomes the separability of two stochastic processes, and *vice versa*.

4.1.3 Analytic Formulation

In section §4.1.1, the ideal filter was defined using the method of *functional analysis*. The definition of an ideal filter can also be derived using an analytic formulation, and such a formulation is readily obtained by introducing a suitable basis onto the signal space $\mathscr{X}$.

Definition 7 (Spectral Response of an Ideal Filter). An ideal filter is defined as an operator which passes, without distortion, all spectral components within a given range, and rejects all others [346, 415]. A linear operator, $\mathscr{L}$, is said to be an ideal filter over the λ-domain for an input space $\mathscr{X}$ if, for any $\mathbf{x}(t) \in \mathscr{X}$, the representation of $\mathbf{y}(t) = \mathscr{L}\,\mathbf{x}(t)$, $\mathbf{Y}(\lambda)$, on the domain λ is given by:

$$\mathbf{Y}(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin \Lambda_H \\ \mathbf{X}(\lambda) & \text{if } \lambda \in \Lambda_H \end{cases} \tag{4.5}$$

where $\mathbf{X}(\lambda)$ is the representation of $\mathbf{x}(t)$ on the λ-domain, and the acceptance region of $\mathscr{L}$, $\Lambda_H \subset \Lambda$, is a subset of the generalised frequency space Λ. ♢

This definition is slightly less general than Definition 6 since it has characterised the composition of the signal classes. Definition 7 may be expressed in terms of the basis kernels of the domain under consideration by the following theorem:

Theorem 13 (Existence of an Ideal Filter) *A linear operator, $\mathscr{L}$, is said to be an ideal filter if, and only if, there exists a basis, $K(t, \lambda)$, and reciprocal basis, $k(\lambda, t)$, defining the transform pair* (3.8)*, and satisfying* (3.10)*, such that:*

$$\mathscr{L}\{k(\lambda, t)\} = \begin{cases} 0 & \textit{if } \lambda \notin \Lambda_H \\ k(\lambda, t) & \textit{if } \lambda \in \Lambda_H \end{cases} \tag{4.6}$$

PROOF. **Sufficiency**

Let $\mathscr{L}$ be a linear operator satisfying (4.6) with respect to the basis set $k(\lambda, t)$. Then, substituting (4.6) into (3.42c) and using (3.10b), gives:

$$H(\lambda, \hat{\lambda}) = \begin{cases} 0 & \text{if } \lambda \notin \Lambda_H \\ \delta(\lambda - \hat{\lambda}) & \text{if } \lambda \in \Lambda_H \end{cases} \tag{4.7}$$

which, after substitution into (3.43), gives Definition 7. Hence, $\mathscr{L}$ is an ideal filter.

Necessity

Let $\mathscr{L}$ be an ideal filter over the domain of $\lambda \in \Lambda$ for an input space $\mathscr{X}$. Then, from equation (3.43), and by Definition 7, $H(\lambda, \hat{\lambda})$ must satisfy equation (4.7) and, consequently, from equations (3.10b) and (3.42c), $\mathscr{L}\{k(\lambda, t)\}$ is given by (4.6). □

Theorem 14 (Impulse Response of an Ideal Filter) *An ideal filter over the domain*

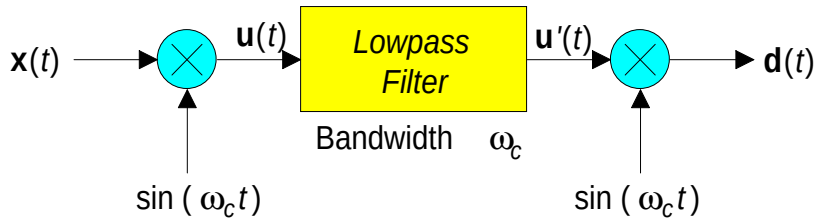


Figure 4.1: *Quadrature demodulator.*

$\lambda \in \Lambda$ *with impulse response* $h(t,\tau)$ *may be represented in the 'separable' form*

$$h(t,\tau) = \int_{\Lambda_H} k(\lambda,t)\, K(\tau,\lambda)\, d\lambda, \quad (t,\tau) \in \mathbf{T} \tag{4.8}$$

where $k(\lambda,t)$ *and* $K(t,\lambda)$ *are related by* (3.10)*, and* Λ_H *is the filter's passband.*

PROOF. Follows from (3.41) and Theorem 13. □

Example 5 (Allpass Ideal Filter). If the acceptance region, Λ_H, is the whole of the signal space, Λ, then the filter must let all signals through. If $\Lambda_H = \Lambda$ then, by comparing the form of equation (4.8) in Theorem 14 with equation (3.10b), and using the superposition integral in (3.35), $y(t) = x(t)$ as required. ⋈

Example 6 (Ideal Filter in Quadrature Modulation). Consider the two quadrature modulated signals:

$$\left.\begin{aligned} \mathbf{d}(t) &= \sin \omega_c t\, \mathbf{a}(t) \\ \mathbf{n}(t) &= \cos \omega_c t\, \mathbf{b}(t) \end{aligned}\right\} \quad t \in \mathscr{T} \subset \mathbb{R} \tag{4.9}$$

where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are bandlimited stochastic signals with Fourier spectral components in the range $(-\omega_c,\ \omega_c)$; $\mathbf{d}(t)$ may be recovered using the quadrature demodulator in Figure 4.1. A similar demodulator is obtained if it is desired to recover $\mathbf{n}(t)$. Analysing this filter by the classical approach of applying an impulse $x(t) = \delta(t-\tau)$ to the input at time τ, the LTV impulse response at time t at the output of the first multiplier is:

$$u(t,\tau) = \sin \omega_c t\, \delta(t-\tau) \tag{4.10}$$

The output of the low pass filter with bandwdith ω_c, which has a LTI impulse

response $h_{LP}(t)$, is given by the LTI convolution:

$$u'(t,\tau) = \int_T h_{LP}(\alpha)\, u(t-\alpha)\, d\alpha \equiv h_{LP}(t-\tau)\ \sin\omega_c\tau$$

Therefore, since the impulse response, $h_{LP}(t)$, of an ideal bandpass filter is given by the 'sinc' function, the LTV impulse response of the complete filter, $h(t,\tau) = u'(t,\tau)\ \sin\omega_c t$, may be written as:

$$h(t,\tau) = 2\frac{\sin\omega_c\,(t-\tau)}{\omega_c\,(t-\tau)}\sin\omega_c t\ \sin\omega_c\tau \tag{4.11}$$

Since $h(t,\tau)$ in (4.11) passes $\mathbf{d}(t)$ without distortion, but rejects $\mathbf{n}(t)$, it is by definition an ideal filter. Note that $h(t,\tau)$ can also be written in separable form, by writing:

$$h(t,\tau) = \frac{1}{\omega_c}\int_{-\omega_c}^{\omega_c} \left(\sin\omega_c t\ e^{j\omega t}\right)\left(\sin\omega_c\tau\ e^{-j\omega\tau}\right)\ d\omega \tag{4.12}$$

where equation (4.12) is in the separable form of equation (4.8). ⋈

4.1.4 Ideal Filters in Compatible Domains

In section §3.3.5 the concept of a compatible transform was discussed. The use of compatible transforms was demonstrated to be difficult, except in some special cases where a system is known to be composed of fundamental elements which possess a compatible transform. Nevertheless, Theorem 13 can be reformulated by considering the form of an ideal filter in a suitable compatible domain. In section §3.3.5.5, it was shown that to yield an algebraic equation, the transform kernel must satisfy

$$\mathscr{L}^* K(t,\lambda) = H(\lambda)\, K(t,\lambda) \tag{3.69b}$$

where $\mathscr{L}^*$ is the adjoint operator to the linear operator $\mathscr{L}$. Suppose that the signal, $x(t)$, is filtered by the time-varying filter, $h(t,\tau)$, such that:

$$\mathscr{L}\, x(t) = \int_T h(t,\tau)\, x(t)\, dt \tag{3.35}$$

The adjoint operator $\mathscr{L}^*$ to the system $\mathscr{L}$ must satisfy

$$\langle \int_T h(t,\tau)\, x(\tau)\, d\tau, K(t,\lambda)\rangle = \langle x(t), \mathscr{L}^* K(t,\lambda)\rangle$$

or,
$$\iint_{\mathbf{T}} h(t,\tau)\, x(\tau) K(t,\lambda)\, d\tau\, dt = \int_T x(t)\, \mathscr{L}^* K(t,\lambda)\, dt \tag{4.13a}$$

Relabelling the variables of integration on the LHS, and interchanging the order of integration gives:

$$\iint_{\mathbf{T}} h(\tau,t)\, x(t)\, K(\tau,\lambda)\, dt\, d\tau = \int_T x(t) \int_T h(\tau,t)\, K(\tau,\lambda)\, d\tau\, dt \tag{4.13b}$$

and thus, comparing the RHS of (4.13b) with (4.13a), $\mathscr{L}^*$ is given by [84, pp. 309]

$$\mathscr{L}^* K(t,\lambda) = \int_T h(\tau,t)\, K(\tau,\lambda)\, d\tau \tag{4.13c}$$

If follows that the compatible transform with kernel $K(t,\lambda)$, for a LTV system with input–output relation given by the superposition integral (3.35), must satisfy:

$$\int_T h(\tau,t) K(\tau,\lambda)\, d\tau = H(\lambda)\, K(t,\lambda) \tag{4.14}$$

where $H(\lambda)$ is some rational function of λ. Hence, the kernel $K(\tau,\lambda)$ are the eigenfunctions of the adjoint linear operator. This result can be used to provide another viewpoint of the sufficiency of Theorem 14:

Theorem 15 (Theorem 14 Revisited) *The sufficiency of Theorem 14 can be proved by reformulating the ideal filter in a compatible domain.*

PROOF. For perfect separation, it is known that an ideal filter is required, and the form of the linear impulse response of a LTV ideal filter is given by:

$$h(t,\tau) = \int_{\Lambda_H} k(\lambda,t)\, K(\tau,\lambda)\, d\lambda, \quad (t,\tau) \in \mathbf{T} \tag{4.8}$$

Substituting (4.8) into equation (4.14) gives:

$$H(\lambda)\, K(t,\lambda) = \int_T \int_{\Lambda_H} k(\hat{\lambda},\tau) K(t,\hat{\lambda}) K(\tau,\lambda)\, d\hat{\lambda}\, d\tau$$

$$
\begin{aligned}
&= \int_{\Lambda_H} K(t,\hat{\lambda}) \int_T k(\hat{\lambda},\tau) K(\tau,\lambda)\, d\tau\, d\hat{\lambda} \\
&= \int_{\Lambda_H} K(t,\hat{\lambda})\, \delta(\hat{\lambda}-\lambda)\, d\hat{\lambda} \\
&= \begin{cases} K(t,\lambda) & \text{if } \lambda \in \Lambda_H \\ 0 & \text{if } \lambda \notin \Lambda_H \end{cases}
\end{aligned}
$$

which, since this must be true for all $t \in T$, simplifies to

$$
H(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \Lambda_H \\ 0 & \text{if } \lambda \notin \Lambda_H \end{cases} \tag{4.15}
$$

which is the definition of the ideal filter. □

4.1.5 Ideal Filter in Discrete Time

Theorem 16 (Discrete-Time Ideal Filter) *An ideal filter in discrete time over the domain $p \in \mathscr{P} \subset \mathbb{Z}$ with response, $h(n,m)$, may be represented in 'separable' form:*

$$
h(n,m) = \sum_{p \in \mathscr{P}_H} k(p,n)\, K(m,p) \tag{4.16}
$$

where $k(p,n)$ and $K(n,p)$ are related by (3.15)*, and $\mathscr{P}_H$ is the acceptance region.*

Using equation (3.14), $\mathbf{X} = K^T \mathbf{x}$, if the ideal filter accepts all spectral components $p \in \mathscr{P}_H$, and rejects all others, the spectral output of the filter may be written as:

$$
\mathbf{Y} = H K^T \mathbf{x} \tag{4.17}
$$

where H is a diagonal matrix with 'ones' in the diagonal elements, corresponding to the spectral components which are to passed without distortion, and zero elsewhere. Hence, noting equations (3.14) and (3.17):

$$
\mathbf{y} = k^T \mathbf{Y} = k^T H K^T \mathbf{x} = K^{-T} H K^T \mathbf{x}
$$

which may be written in the form:

$$
\mathbf{y} = A\,\mathbf{x} \quad \text{where} \quad A = K^{-T} H K^T \tag{4.18a}
$$

Now consider, for clarity, the instance when the pass-band is contiguous[1] and contains the first N spectral components: then

$$H = \left[\begin{array}{c|c} I_N & 0 \\ \hline 0 & 0 \end{array}\right] \equiv \left[\begin{array}{c|c} I_N & 0 \\ \hline 0 & M \end{array}\right] \tag{4.18b}$$

where M is an arbitrary matrix and, hence, A may be decomposed as:

$$A = k_d{}^T K_d{}^T + M\, k_v{}^T K_v{}^T \tag{4.18c}$$

where $K_d{}^T$ and k_d are the appropriately defined basis matrices for the pass-band components, and $K_v{}^T$ and $k_v{}^T$ are the basis matrices for the spectral components of the unused space. Equation (4.18c) may also be obtained by extending the basis in the transform domain so as to span the whole vector space, *e.g.* see [347].

4.1.5.1 Noise Gain of Ideal Filter

Suppose an ideal filter $h(t,\tau)$ is constructed to separate two signals over a domain λ with passband Λ_H. If an additional noise signal, $\mathbf{w}(t)$, appears at the input of the filter then, in general, it is passed to the filter's output. The *noise gain* of an ideal filter is defined as the gain in noise energy when white noise passes through it.

Theorem 17 (Noise Gain) *The noise gain of an ideal filter over the domain $\lambda \in \Lambda$ with passband Λ_H is given by:*

$$N_{WGN} = \iint_{\Lambda_H} \bar{k}(\lambda,\hat{\lambda})\, d\lambda\, d\hat{\lambda} \tag{4.19a}$$

where

$$\bar{k}(\lambda,\hat{\lambda}) \triangleq \int_T k(\lambda,t)\, k^*(\hat{\lambda},t)\, dt \tag{4.19b}$$

where $k(\lambda,t)$ and its inverse, $K(t,\lambda)$, are the basis functions defining the transform domain, and are related by equations (3.10)*. If $k(t,\lambda)$ and $K(t,\lambda)$ are self-reciprocal, then the noise gain simplifies to:*

$$N'_{WGN} = \Lambda_H \leq \Lambda \tag{4.20}$$

PROOF. The output of the filter $h(t,\tau)$ at time t, when the excitation is a WGN

[1] If they are not, then a permutation matrix can be used to ensure that they are contiguous.

sequence $\{w(t)\}$, is given by the superposition integral equation (3.35):

$$\mathbf{y}(t) = \int_T h(t,\tau)\,\mathbf{w}(\tau)\,d\tau, \quad \forall t \in T \tag{3.35}$$

and, therefore, the variance of the noise at the output is:

$$\sigma^2_{yy}(t) = \sigma^2_{ww}\int_T |h(t,\tau)|^2\,d\tau$$

Substituting the spectral expansion of the ideal filter using equation (4.8) gives, on expansion, the desired result. If $k(t,\lambda)$ and $K(t,\lambda)$ are self-reciprocal then $k(t,\lambda) = K^*(t,\lambda)$ and, using equation (3.10b), equation (4.19b) reduces to (4.20). □

4.2 General Spectral Solution of the Nonstationary Wiener-Hopf Filter

General solutions to the nonstationary Wiener-Hopf filter defined in equation (3.5a) of Theorem 1 are not easily found, although it is easy to show, as for instance in Miller and Zadeh [251], that the solution of the Wiener-Hopf equations can be reduced to the factorisation of the autocorrelation functions. For example, Miller and Zadeh [251] discuss the factorisation of the autocorrelation function in terms of its innovation filter $G(t,\omega)$ defined in equation (3.22). Further, the methods of Laning and Battin [207], Zadeh and Miller [251], Shinbrot [324, 325], Darlington [85] and others, where further references may be found in the classic review paper by Zadeh [417], for solving equation (3.5a) are either explicitly based on the factorisation of $R_{xx}(t,\tau)$ and $R_{dx}(t,\tau)$, or make implicit use of it: *e.g.*, Shinbrot [324, 325] and Darlington [85] assume the innovations filter satisfies a LTV differential equation, such that the autocorrelation functions are of the form:

$$R(t,\tau) = \begin{cases} \sum_{i=1}^{N} \phi_i(t)\,\gamma_i(\tau), & t > \tau \\ \sum_{i=1}^{N} \phi_i(\tau)\,\gamma_i(t), & t \le \tau \end{cases} \tag{4.21}$$

Similar considerations are made by Youla [407] when considering solutions to the stationary Wiener-Hopf filter equations. It is also of interest to note that some techniques which attempt to estimate the auto- and cross-correlation functions of nonstationary processes from a single realisation rely on a similar decomposition (see section §4.3 and [110, 242, 243]).

In [156, 192, 196], nonstationary Wiener filter solutions are presented in terms of the Wigner-Ville time-frequency decomposition [75] and, recently, Khan and Chaparro [190] present a solution in terms of *evolutionary spectrum* [295, 296]. Here, a sufficient solution for the nonstationary Wiener-Hopf filter is presented which relies on the factorisation of the autocorrelation function into the form of the generalised power spectrum introduced in section §3.2. In the signal separability problem, it is desirable to find a filter solution which also satisfies equation (3.5c) for perfect signal separation. The details of the solution for the stationary Wiener filter in the independent additive case (see Definition 3 on page 37) are investigated in Appendix C.3, and this provides an interesting foundation for the form of the solution to the nonstationary Wiener-Hopf filter, as discussed in the section §4.2.2.

4.2.1 Physical Realisability

Generally, an ideal filter is not physically realisable since its impulse response, $h_{ideal}(t, \tau)$, does not necessarily vanish for $t < \tau$, and is therefore noncausal. If it is desirable that the solution of the Wiener filter, $h(t, \tau)$, should vanish for $t < \tau$, an extra constraint will be added onto the separability criterion and the *spectral solution* of the Wiener filter, in the general case, becomes more complicated and less intuitive. However, following the argument in Bode and Shannon, if $h(t, \tau)$ is noncausal then it can can be approximated by a physically realisable system of impulse response $h(t - t_0, \tau)$, where t_0 is a sufficiently large time delay, provided that $h(t, \tau) \to 0$ as $t - \tau \to \infty$. The constraint of physical realisability is left for investigation in subsequent research since, in the context of this work, it is desirable to use *all possible prior* knowledge of a system to determine separability. It follows in the nonstationary problem that future values of the ACFs must be used, as well as past ones and, as such, a noncausal filter is assumed. Abdrabbo and Priestley [1, 2] investigate the prediction and filtering problem for nonstationary processes using the evolutionary spectrum using the constraint of physical realisability.

4.2.2 Spectral Solution

A sufficient solution to the Wiener filter equations can be readily obtained by the factorisation of the autocorrelation functions into the GPSs. A necessary condition for a solution is not presented, inasmuch as there may be solutions for signal structures other than that defined in equation (3.18); this result is presented in [162].

Theorem 18 (Wiener-Hopf Filter Solution) *Suppose* $R_{dd}(t,\tau)$, $R_{nn}(t,\tau)$ *and* $R_{dn}(t,\tau)$ *can be written as the generalised spectral decompositions:*

$$R_{dd}(t,\tau) = \iint_{\hat{\Lambda}_d \hat{\Lambda}_d} \mathcal{P}_{dd}(\lambda,\hat{\lambda})\, k(\lambda,t)\, k^*(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.22a}$$

$$R_{nn}(t,\tau) = \iint_{\hat{\Lambda}_n \hat{\Lambda}_n} \mathcal{P}_{nn}(\lambda,\hat{\lambda})\, k(\lambda,t)\, k^*(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.22b}$$

$$R_{dn}(t,\tau) = \iint_{\boldsymbol{\Lambda}_0} \mathcal{P}_{dn}(\lambda,\hat{\lambda})\, k(\lambda,t)\, k^*(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.22c}$$

where $(t,\tau) \in \mathbf{T}$, $\hat{\Lambda}_d \equiv \Lambda_d \oplus \Lambda_0$ *and* $\hat{\Lambda}_n \equiv \Lambda_n \oplus \Lambda_0$. *Here,* $\Lambda_d \subset \Lambda$ *and* $\Lambda_n \subset \Lambda$ *are the regions of the* $\Lambda \subset \mathbb{C}$ *space over which the spectral components of* $\{\mathbf{d}(t),\ t \in T\}$ *and* $\{\mathbf{n}(t),\ t \in T\}$, *respectively, do* not *overlap, and* $\Lambda_0 \subset \Lambda$ *is the region over which spectral components of* $\mathbf{d}(t)$ *and* $\mathbf{n}(t)$ do *overlap. Hence,* $\Lambda_d \cap \Lambda_n = \{\emptyset\}$, $\Lambda_d \cap \Lambda_0 = \{\emptyset\}$ *and* $\Lambda_n \cap \Lambda_0 = \{\emptyset\}$. *A sufficient solution,* $h(t,\tau)$, *to the WHF equation* (3.5a) *for the additive case,*[2] *when* $\mathcal{T} \equiv T \subset \mathbb{R}$, *is given by equation* (3.38a):

$$h(t,\tau) = \iint_{\boldsymbol{\Lambda}} H(\lambda,\hat{\lambda})\, k(\lambda,t)\, K(\tau,\hat{\lambda})\, d\lambda\, d\hat{\lambda} \tag{3.38a}$$

where $k(\lambda,t)$ *and* $K(t,\lambda)$ *are related by* (3.10), *and the GBTF* $H(\lambda,\hat{\lambda})$, *is given by:*

$$H(\lambda,\hat{\lambda}) = \begin{cases} \delta(\lambda-\hat{\lambda}) & \text{for } \{\lambda,\hat{\lambda}\} \in \boldsymbol{\Lambda}_d, \\ H_0(\lambda,\hat{\lambda}) & \text{for } \{\lambda,\hat{\lambda}\} \in \boldsymbol{\Lambda}_0, \\ 0 & \text{elsewhere.} \end{cases} \tag{4.23a}$$

where $\boldsymbol{\Lambda}_d = \Lambda_d \times \Lambda_d$, $\boldsymbol{\Lambda}_0 = \Lambda_0 \times \Lambda_0$, *and the function* $H_0(\lambda,\hat{\lambda})$ *is the solution of*

$$\mathcal{P}_{dx}(\lambda,\hat{\lambda}) = \int_{\Lambda_0} H_0(\lambda,\bar{\lambda})\, \mathcal{P}_{xx}(\bar{\lambda},\hat{\lambda})\, d\bar{\lambda},\ \forall(\lambda,\hat{\lambda}) \in \boldsymbol{\Lambda}_0 \tag{4.23b}$$

[2] See Definition 3 on page 37.

Further, $H_0(\lambda,\hat{\lambda}) \triangleq 0$, $\forall\,(\lambda,\hat{\lambda}) \notin \boldsymbol{\Lambda}_0$. *The filter,* $h(t,\tau)$, *may, therefore, be written as*

$$h(t,\tau) = \underbrace{\int_{\Lambda_d} k(\lambda,t)\,K(\tau,\lambda)\,d\lambda}_{h_{\text{ideal}}(t,\tau)} + \underbrace{\iint_{\boldsymbol{\Lambda}_0} H_0(\lambda,\hat{\lambda})\,k(\lambda,t)\,K(\tau,\hat{\lambda})\,d\lambda\,d\hat{\lambda}}_{h_{\text{overlap}}(t,\tau)} \tag{4.23c}$$

where $\mathcal{P}_{xx}(\cdot) = \mathcal{P}_{dx}(\cdot) + \mathcal{P}_{nx}(\cdot)$, $\mathcal{P}_{dx}(\cdot) = \mathcal{P}_{dd}(\cdot) + \mathcal{P}_{dn}(\cdot)$ *and* $\mathcal{P}_{nx}(\cdot) = \mathcal{P}_{nd}(\cdot) + \mathcal{P}_{nn}(\cdot)$. *The resulting mean squared error (MSE) is given by:*

$$\sigma^2(t) = \iint_{\boldsymbol{\Lambda}_0} \mathcal{P}_{\sigma\sigma}(\lambda,\hat{\lambda})\,k(\lambda,t)\,k^*(\hat{\lambda},t)\,d\lambda\,d\hat{\lambda} \tag{4.23d}$$

where the spectrum of $\sigma^2(t)$ *is given by:*

$$\mathcal{P}_{\sigma\sigma}(\lambda,\hat{\lambda}) = \mathcal{P}_{dd}(\lambda,\hat{\lambda}) - \int_{\Lambda_0} H_0(\lambda,\bar{\lambda})\,\mathcal{P}_{dx}(\bar{\lambda},\hat{\lambda})\,d\bar{\lambda} \tag{4.23e}$$

$$= \int_{\Lambda_0} H_0(\lambda,\bar{\lambda})\left\{\mathcal{P}_{nd}(\bar{\lambda},\hat{\lambda}) + \mathcal{P}_{nn}(\bar{\lambda},\hat{\lambda})\right\} d\bar{\lambda} \tag{4.23f}$$

It is assumed the initial state of the filter $h(t,\tau)$ *is at* rest.

PROOF. The proof is given in Appendix C.4 and is essentially by substitution. □

Note the similarity of equation (4.23b) to equation (3.5a) due to the change in basis. However, the simplification in equation (4.23b) has occurred since, in the λ-domain, the signals are assumed to have non-overlapping components.

4.2.2.1 Ideal Filter Component

Following from Theorem 14, the first term in (4.23c) corresponds to the expression for an ideal filter, $h_{\text{ideal}}(t,\tau)$, which passes the frequency range Λ_d and thus corresponds to the perfect separation of the non-overlapping signal components. Moreover, the first term in (4.23c), unlike the second, does not depend on *full* knowledge of the signals autocorrelation functions, inasmuch as it does not depend on the actual values of the signals $\mathbf{n}(t)$ or $\mathbf{d}(t)$, but only on the 'generalised' frequency bands over which the signal components do not overlap. As such, a condition for perfect signal separation is immediately obtained:

Theorem 19 (Ideal Filter Component of Wiener Solution) *In the additive case, as in Definition 3, the solution to the Wiener-Hopf equations* (3.5a) *and* (3.5c), $h(t,\tau)$,

which gives perfect separation, is an ideal filter, assuming that the signals have disjoint spectral components.

PROOF. This is a direct corollary of Theorem 18. The mean squared error (4.23d) is zero, if the region over which the spectral components do overlap, $\Lambda_0 \equiv \{\emptyset\}$. This implies that the desired and noise signals are independent processes. In such an instance, the Wiener filter consists of the first term in (4.23c), and has the form of an ideal filter given by equation (4.8). □

Note, by Theorem 19, that for signals to be perfectly separable, they are required to be spectrally disjoint and, therefore, bandlimited in that domain.[3]

The idea of signals being separable if they are spectrally disjoint is not new; for separating periodic signals, it is the basis of methods such as *harmonic selection* [287], nonlinear *Cepstrum filtering* [340–342], algebraic techniques and comb filters [315, 316, 420]. As well as being analogous to the result in the stationary case, Abdrabbo [1] notes that in the case of the evolutionary spectrum, 'it is impossible to recover $\mathbf{d}(t)$ with perfect precision unless the signal and noise have non-overlapping evolutionary spectra.' Indeed, the result derived in Theorem 19 for the general nonstationary case is as intuitively appealing, from a filtering perspective, as it is in the stationary case. Note that since the ideal filter is independent of the signal values, it separates the class of all nonstationary stochastic signals that are disjoint in the filter's domain, provided the desired signal components lie in its passband. The only *prior knowledge* required for separation is the *domain* in which the signals are disjoint, and the *spectral regions* over which the spectral components lie. This means that the prior knowledge is kept to a minimum, and knowledge of the full autocorrelation function is not required.

4.2.2.2 Signal Dependent Component and Resulting MSE

The second term in equation (4.23c) is signal dependent, given by the solution of (4.23b). As expected, the filtered power spectrum of each process in the overlapping spectral region sums to the power spectra of the desired signal. The MSE is the energy of the filtered noise process contained in the overlapping spectra.

[3] It is of interest to note the result by Papoulis [285], which states that the values of a sampled Fourier bandlimited process $\{\mathbf{x}(t),\, t \in \mathbb{Z}\}$ are linearly dependent: $y(t) = \sum_{\tau=-\infty}^{\infty} c(\tau)\, x(t-\tau) = 0$ for a particular set of coefficients $c(\tau)$. This result can be extended to a general transform domain.

4.2.3 Signal Separability

Theorems 18 and 19 lead to conditions that are placed on a signal to obtain separability to a given accuracy, and lead to the following criterion:

Theorem 20 (Perfect Single Channel Signal Separation) *Perfect single channel signal separation is only possible if there exists some domain where the generalised spectral representations of the desired and noise signals are disjoint. Thus, the desired and noise signals must be uncorrelated.*

4.3 Estimation of Correlation Functions

So far, emphasis has been placed on the fact that the autocorrelation functions of nonstationary random processes are rarely known in full and, as such, a theory has been developed for signal separation which relies on only partial knowledge of the correlation function. If the correlation functions $R_{dd}(t,\tau)$, $R_{xx}(t,\tau)$ and $R_{dx}(t,\tau)$ are known in full, single channel signal separation using LTV filters reduces to the task of finding the optimal solution of the Wiener filter equations (3.5a) and (3.5c), with $h(t,\tau)$ as the only unknown. In practice, obtaining *ensemble* averages is an extremely burdensome task, and often impossible. However, using the decompositions of the autocorrelation function and power spectrum similar to those introduced in section §3.2, some time-average estimators have been proposed which yield unbiased and consistent estimates of such correlation functions from a single record of the underlying random process, thus waiving the limitations imposed by the *ergodic hypothesis* [286, 335]. These estimators are discussed in Snyder [332], Marmarelis [242,243], Boyles and Gardner [50,110]; Wu and Lev-Ari [399] discuss other forms of nonstationary moment estimators. The single-record time-averaged correlation estimators of a random process proposed by Marmarelis [242,243] are briefly discussed below to provide an approach for single channel signal separation using the Wiener-Hopf filter where the autocorrelation functions are estimated from the data. In this time-average estimator, suppose the correlation function $R_{dx}(t,\tau) = E[\mathbf{d}(t)\,\mathbf{x}^*(\tau)]$ can be written in the form

$$R_{dx}(t,\tau) \triangleq \sum_{m=-M}^{M} a_m(\tau)\, b_m(t) \tag{4.24a}$$

where $\mathscr{B}(t) = \{b_m(t),\ m \in \mathscr{M}\}$, $\mathscr{M} = \{-M, \ldots, M\}$, is a complete set of orthonormal functions on the interval $[0, T]$, known *a priori* such that

$$\frac{1}{T}\int_0^T b_m(t)\, b_n^*(t)\, dt = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \tag{4.24b}$$

and $\{a_m(\tau),\ m \in \mathscr{M}\}$ is a set of τ-dependent coefficients. An estimator for $R_{dx}(t, \tau)$ is given in terms of $\mathscr{B}(t)$ by:

$$\hat{R}_{dx}(t, \tau) \triangleq \sum_{m=-M}^{M} \hat{a}_m(\tau)\, b_m(t) \tag{4.24c}$$

where

$$\hat{a}_m(\tau) \triangleq \frac{1}{T}\int_0^T d(t)\, x(\tau)\, b_m^*(t)\, dt \tag{4.24d}$$

Marmarelis [242, 243] shows that this estimator yields unbiased and consistent estimators of the correlation function of the underlying process.

4.4 Selecting Transform Kernels

In section §4.2, it is demonstrated that if two signals lie in some domain, λ, and are virtually disjoint, then *perfect signal separation* is achievable given appropriate prior knowledge of the signal structure. However, no mention has yet been indicated of how the transform domains are chosen, except in the case of stationary signals. One valid solution to the signal separability problem is to choose all possible kernels, $\{k(\lambda, t),\ (\lambda, t) \in \Lambda \times T\}$, and classify as separable all pairs of stochastic processes admitted by this representation that are disjoint in this domain. Such an approach is exhaustive and does not provide a satisfactory solution to the more general question of whether two processes are separable given the form of their autocorrelation functions. Although a general approach to answering this question is difficult, and remains elusive, the following work is presented to give some insight into answering the question. The first contribution is based on the theory of *compatible transforms*, as discussed in section §3.3.5. The second, and more powerful, contribution discussed in the next section relies on finding a common domain in which the representation of two signals are disjoint. This is achieved by *concatenating* the basis functions for each individual signal representation to find

a common domain. To help simplify this task, it is noted that the kernel functions, $k(\lambda, t)$ and $K(t, \lambda)$, used in the stochastic spectral representation of a stochastic process, as given by equations (3.18), are, by design, the same kernels used in the power spectral representation of the ACF as given in Definition 4. This fact is useful since the kernel functions of the stochastic spectrum arise more naturally from signal structures than from the structure of the ACF. Thus, if $k(\lambda, t)$ and $K(t, \lambda)$ can be deduced from the signal structure, they can be directly substituted as the kernel functions in the generalised power spectrum.

4.4.1 Concatenating Power Spectra

The characteristics of the Wiener-Hopf filter in Theorem 18 were derived on the assumption that the spectral representations of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are defined on a common domain. This, however, appears restrictive and seems to indicate, for example, that a stationary process can be separated from a nonstationary process if, and only if, they are disjoint in the Fourier domain (see Theorem 7 on page 50); the discussion in Chapter 1 clearly demonstrates this is not the case. This apparent restriction is lifted by concatenating the different spectral representations of each signal, $\mathbf{d}(t)$ and $\mathbf{n}(t)$, provided each representation is bandlimited in their respective domains.[4] So, for example, suppose $\mathbf{d}(t)$ is a stationary process expressed in the Fourier domain, $\omega \in \Omega \subset \mathbb{R}$, with basis set $\{k_f(\omega, t) = e^{-j\omega t},\ (\omega, t) \in \Omega \times T\}$, and $\mathbf{n}(t)$ is a nonstationary process, where $\mathbf{d}(t)$ and $\mathbf{n}(t)$ overlap in the Fourier domain, thus avoiding trivial cases. If $\mathbf{n}(t)$ is expressed in a domain, $\lambda \in \Lambda$, with basis set $\{k_n(\lambda, t),\ (\lambda, t) \in \Lambda \times T\}$, where its representation $\mathbf{N}(\lambda)$ is bandlimited, these domains can be combined as follows: the kernel in this *concatenated* domain, $\hat{\lambda} \in \hat{\Lambda} \subset \mathbb{C}$, denoted by $\{k(\hat{\lambda}, t),\ (\hat{\lambda}, t) \in \hat{\Lambda} \times T\}$, is constructed as:

$$k(\hat{\lambda}, t) = \begin{cases} k_n(\lambda(\hat{\lambda}), t) & , \quad \{\hat{\lambda} : |\mathbf{N}(\lambda(\hat{\lambda}))| \geq 0 \quad , \lambda \in \Lambda \subset \mathbb{C}\} \\ k_f(\omega(\hat{\lambda}), t) & , \quad \{\hat{\lambda} : |\mathbf{D}(\omega(\hat{\lambda}))| \geq 0 \quad , \omega \in \Omega \subset \mathbb{R}\} \end{cases} \tag{4.25}$$

where $\lambda(\hat{\lambda})$ and $\omega(\hat{\lambda})$ are isomorphic mappings of $\lambda : \hat{\Lambda} \to \Lambda$ and $\omega : \hat{\Lambda} \to \Omega$, with $\hat{\Lambda} \subset \mathbb{C}$. In other words, the kernel in the region of $\hat{\Lambda}$ on which the stationary signal is represented corresponds to the Fourier kernel, while the kernel in the region of $\hat{\Lambda}$ on which the nonstationary signal is represented corresponds to $k_n(\lambda(\hat{\lambda}), t)$. Assuming these regions are disjoint, this domain exists if the transform is unique

[4] Note that if a signal has infinite bandwidth in one domain, it may (and is likely to) have finite bandwidth in some other domain; it is assumed that this domain has been identified.

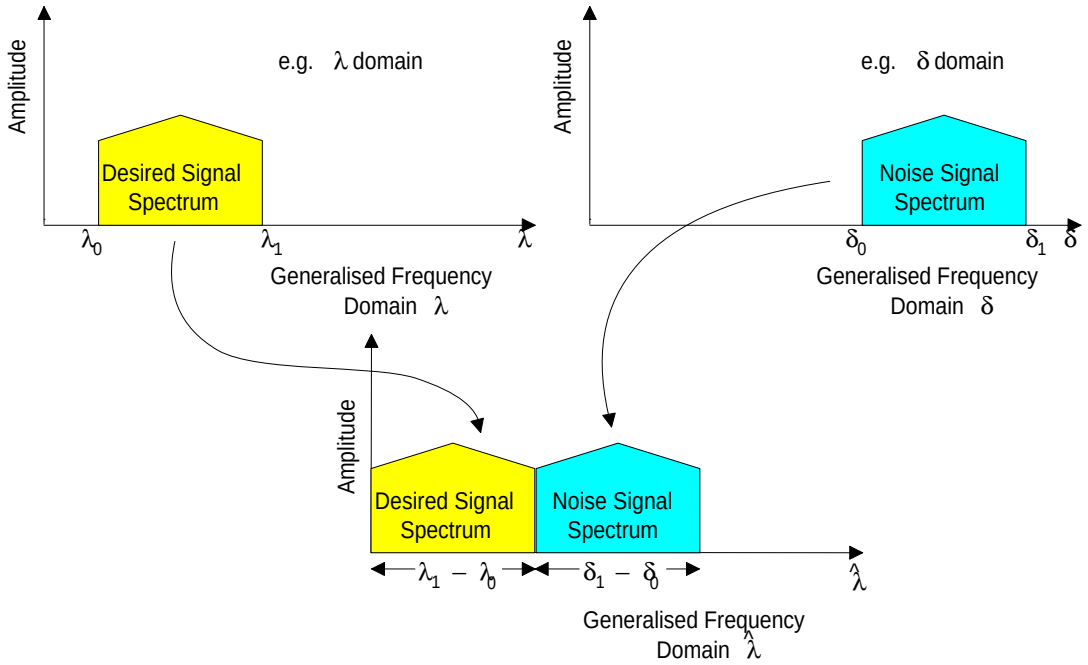


Figure 4.2: *Concatenating power spectra by changing the variables of the generalised frequency axis and representing both signals on a common domain.*

and, in such a case, the two signal classes are separable. If the signals overlap in the $\hat{\lambda}$-domain, then a transform pair cannot be derived since there is no unique transform. The idea behind this 'concatenation' is summarised in Figure 4.2.

To reiterate this methodology, suppose that the autocorrelation functions of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ admit the spectral representations:

$$R_{dd}(t,\tau) = \iint_{\boldsymbol{\Lambda}_D} \mathcal{P}_{dd}(\lambda,\hat{\lambda})\, k_d(\lambda,t)\, k_d^*(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.26a}$$

$$R_{nn}(t,\tau) = \iint_{\boldsymbol{\Lambda}_N} \mathcal{P}_{nn}(\lambda,\hat{\lambda})\, k_n(\lambda,t)\, k_n^*(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.26b}$$

for any $\Lambda_d \subset \Lambda$ and $\Lambda_n \subset \Lambda$. Now, assume it is possible to apply a change of variables $\lambda'_d : \Lambda_d \to \Lambda'_d$ and $\lambda'_n : \Lambda_n \to \Lambda'_n$ such that equations (4.26) become:

$$R_{dd}(t,\tau) = \iint_{\boldsymbol{\Lambda}'_D} \mathcal{P}'_{dd}(\lambda,\hat{\lambda})\, k'_d(\lambda,t)\, k'^*_d(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.27a}$$

$$R_{nn}(t,\tau) = \iint_{\boldsymbol{\Lambda}'_N} \mathcal{P}'_{nn}(\lambda,\hat{\lambda})\, k'_n(\lambda,t)\, k'^*_n(\hat{\lambda},\tau)\, d\lambda\, d\hat{\lambda} \tag{4.27b}$$

where $\Lambda'_{\mathrm{D}} \cap \Lambda'_{\mathrm{N}} = \{\emptyset\}$. In such a case, the spectral representations can be concatenated by defining the *concatenated kernel*:

$$k(\lambda, t) = \sum_{i} \mathbb{I}_{\Lambda_i}(\lambda)\ k'_i(\lambda, t) \tag{4.28}$$

$\forall(\lambda, t) \in \Lambda \times \mathsf{T}$, $i \in \boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the set of indices for the various spectral subsets, taking on the elements 'd', 'n', and 'v' so $\boldsymbol{\theta} \triangleq \{d, n, v\}$; correspondingly, $\boldsymbol{\Theta} \triangleq \{\mathrm{D}, \mathrm{N}, \mathrm{V}\}$ contains the elements 'D', 'N' and 'V'. The indicator function $\mathbb{I}_{\Lambda}(\lambda)$ is defined as,

$$\mathbb{I}_{\Lambda}(\lambda) = \begin{cases} 1, & \text{if } \lambda \in \Lambda \\ 0, & \text{otherwise} \end{cases}$$

Note further that Λ_{V} is 'unused' space, such that $\bigoplus_{i \in \boldsymbol{\Theta}} \Lambda_i \equiv \Lambda_{\mathrm{D}} \oplus \Lambda_{\mathrm{N}} \oplus \Lambda_{\mathrm{V}} = \Lambda$. Assuming this transform domain exists, it follows that the autocorrelation functions of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ possess spectral representations on a domain with kernel $k(\lambda, t)$ as defined in equation (4.28), rather than the kernels $k'_d(\lambda, t)$ and $k'_n(\lambda, t)$ in equation (4.27). The *forward* transforms for $\mathcal{P}_{dd}(\lambda, \hat{\lambda})$ and $\mathcal{P}_{nn}(\lambda, \hat{\lambda})$ are given by the form of equation (3.24a), with the forward transform kernel, $K(t, \lambda)$, corresponding to $k(\lambda, t)$, satisfying equations (3.10). Since each of the autocorrelation functions have been represented on the same transform domain, assuming one exists, it follows that these processes can be separated by the ideal filter:

$$h(t, \tau) = \int_{\Lambda_{\mathrm{H}}} k(\lambda, t)\, K(\tau, \lambda)\, d\lambda \tag{4.8}$$

Ergo, from section §3.2.2, such a domain exists when $k(\lambda, t)$ and $K(t, \lambda)$ satisfy equations (3.10).

4.4.2 Concatenating Discrete Spectra

Following the same line of argument as in the continuous case, suppose that the autocorrelation functions of the two processes $\mathbf{d}(n)$ and $\mathbf{n}(n)$ can be written as two-dimensional transform pairs with kernels $k_d(p, n)$ and $k_n(p, n)$, respectively, and domains of summation given by $\mathscr{P}_{\mathrm{D}}$ and $\mathscr{P}_{\mathrm{N}}$, respectively, where $\bigoplus_{i \in \boldsymbol{\Theta}} \mathscr{P}_i = \mathscr{P}$, P_{V} is unused space, and $\mathscr{P}_{\mathrm{D}} \cap \mathscr{P}_{\mathrm{N}} = \{\emptyset\}$ *etc.* For such a domain to exist, the concatenated kernel $k(p, n)$ must have an inverse. In the discrete case, $\{k(p, n), (p, n) \in \mathscr{P} \times \mathscr{N}\}$ can be written as a matrix, and thus the existence of

the transform can be found by using linear algebra techniques. Assume, for clarity, that $\{\mathscr{P}_i, i \in \boldsymbol{\Theta}\}$ are contiguous regions,[5] given by $\mathscr{P}_D = \{0, \ldots, P_D - 1\}$, $\mathscr{P}_N = \{P_D, \ldots, P_N - 1\}$, and $\mathscr{P}_V = \{P_N, \cdots, P - 1\}$, where P denotes the number of element in $\mathscr{P}$; *i.e.* $P = \dim \mathscr{P}$. If the matrices k_d, k_n and k_v, whose elements are the values of their respectively named kernels, are appropriately defined then, using the notation of section §3.2.2.2, it can easily be shown that:

$$k\,K = \left[\begin{array}{c} k_d \\ \hline k_n \\ \hline k_v \end{array}\right] \left[\begin{array}{c|c|c} K_d & K_n & K_v \end{array}\right] = I_P \tag{4.29}$$

If separation is possible, this transform must be unique, and the kernels must be transform pairs: thus the square *kernel matrix*, k, must be invertible and, therefore, have full rank. This implies any column of one of these partitioned matrices cannot lie within the column space of the other matrices. It is always possible to extend the basis set with k_v to obtain a full rank matrix if k_d and k_n are linearly independent, since k_d and k_n form a subspace of the complete space.

4.5 Separating Modulated Signals

A method is presented in the previous section for combining the kernel functions of the representations of different signal classes into a single transform. Using this method, separability constraints can be determined for the class of *filtered modulated* signals which exhibit the general form:

$$\left.\begin{aligned} \mathbf{d}(t) &= \int_T h_d(t,\tau)\,\mathbf{a}(\tau)\,d\tau \\ \mathbf{n}(t) &= \int_T h_n(t,\tau)\,\mathbf{b}(\tau)\,d\tau \end{aligned}\right\} \quad t \in T \tag{2.1a}$$

where, assuming some prior knowledge regarding the structure of the signals $\mathbf{d}(t)$ and $\mathbf{n}(t)$, it is known that $\mathbf{d}(t)$ and $\mathbf{n}(t)$ overlap in the time and Fourier spectral domains (to avoid trivial cases), $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are bandlimited to $\pm\omega_c$ but otherwise unknown, and $h_d(t,\tau)$ and $h_n(t,\tau)$ are known deterministic signals. This problem will be solved by *concatenating* the *power spectra* of each of the processes.

[5] If they are not, then a permutation matrix can be used such that they are contiguous.

4.5.1 Continuous Modulation

Since $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are bandlimited in the Fourier domain, they admit the representation

$$\begin{aligned}\mathbf{a}(t) &= \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} \mathbf{A}(\omega)\, e^{j\omega t}\, d\omega \\ \mathbf{b}(t) &= \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} \mathbf{B}(\omega)\, e^{j\omega t}\, d\omega\end{aligned} \tag{4.30a}$$

and, therefore, after substitution into (2.1a), a little rearrangement, and use of the mappings $\lambda_d : \Omega \to \Omega + \omega_c$ for $\mathbf{d}(t)$ and $\lambda_n : \Omega \to \Omega - \omega_c$ for $\mathbf{n}(t)$, it follows that $\mathbf{d}(t)$ and $\mathbf{n}(t)$ admit the representations:

$$\begin{aligned}\mathbf{d}(t) &= \int_{0}^{2\omega_c} \mathbf{D}(\lambda)\, k(\lambda, t)\, d\lambda \\ \mathbf{n}(t) &= \int_{-2\omega_c}^{0} \mathbf{N}(\lambda)\, k(\lambda, t)\, d\lambda\end{aligned} \tag{4.30b}$$

where $\lambda \in \Lambda \subset \mathbb{R}$, and the kernel is defined as:

$$k(\lambda, t) = \begin{cases} \frac{1}{2\pi}\int_T h_d(t,\tau)\, e^{j(\lambda-\omega_c)\tau}\, d\tau & \text{for } \lambda \in [0, 2\omega_c) \\ \frac{1}{2\pi}\int_T h_n(t,\tau)\, e^{j(\lambda+\omega_c)\tau}\, d\tau & \text{for } \lambda \in [-2\omega_c, 0) \\ k_v(\lambda, t) & \text{for } \lambda \notin (-2\omega_c, 2\omega_c) \end{cases} \tag{4.31a}$$

with the generalised spectral components given by $\mathbf{D}(\lambda) = \mathbf{A}(\lambda - \omega_c)$ and $\mathbf{N}(\lambda) = \mathbf{B}(\lambda + \omega_c)$. The form of $k_v(\lambda, t)$ is be chosen to complete a basis function set. It follows that a convenient form of expressing the inverse kernel is:

$$K(t, \lambda) = \begin{cases} \int_T g_d(t,\tau)\, e^{-j(\lambda-\omega_c)\tau}\, d\tau & \text{for } \lambda \in [0, 2\omega_c) \\ \int_T g_n(t,\tau)\, e^{-j(\lambda+\omega_c)\tau}\, d\tau & \text{for } \lambda \in [-2\omega_c, 0) \\ K_v(t, \lambda) & \text{for } \lambda \notin (-2\omega_c, 2\omega_c) \end{cases} \tag{4.31b}$$

If $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are separable, $k(\lambda, t)$ and $K(t, \lambda)$ must satisfy equations (3.10): substitution of equations (4.31) into (3.10) creates constraints on the filters $h_d(t, \tau)$ and $h_n(t, \tau)$. In the general continuous case, finding general constraints on the filters $h_{[\boldsymbol{\theta}]_i}(t, \tau)$ is difficult, and more tractable results are obtained by considering the discrete–time case.

4.5.2 Discrete Modulation

The results from section §4.5.1 carry across, *mutatis mutandis*, to the discrete-time discrete-spectrum case:

$$\mathbf{a}(n) = \frac{1}{N}\sum_{-q_c}^{q_c} \mathbf{A}(q)\, e^{jnq\frac{2\pi}{N}} \quad \text{and} \quad \mathbf{b}(n) = \frac{1}{N}\sum_{-q_c}^{q_c} \mathbf{B}(q)\, e^{jnq\frac{2\pi}{N}} \tag{4.32a}$$

where $\mathbf{a}(n), \mathbf{b}(n), \mathbf{d}(n), \mathbf{n}(n) \in \mathbb{R}^N$, and $\mathbf{a}(n)$ and $\mathbf{b}(n)$ are bandlimited to q_c. Hence, $\mathbf{d}(n)$ and $\mathbf{n}(n)$ admit the representationss:

$$\mathbf{d}(n) = \sum_{p\in\mathscr{P}} \mathbf{D}(p)\, k(p,n) \quad \text{and} \quad \mathbf{n}(n) = \sum_{p\in\mathscr{P}} \mathbf{N}(p)\, k(p,n) \tag{4.32b}$$

where $\mathbf{D}(p) = \mathbf{A}(p - q_c)$, $\mathbf{N}(p) = \mathbf{B}(p + q_c)$, and

$$k(p,n) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{\hat{n}\in\mathscr{N}} h_d(n,\hat{n})\, e^{j\hat{n}(p-q_c)\frac{2\pi}{N}} & \text{for } p \in \mathscr{P}_D \\ \dfrac{1}{N}\displaystyle\sum_{\hat{n}\in\mathscr{N}} h_n(n,\hat{n})\, e^{j\hat{n}(p+q_c)\frac{2\pi}{N}} & \text{for } p \in \mathscr{P}_N \\ k_v(p,n) & \text{for } p \in \mathscr{P}_V \end{cases} \tag{4.33a}$$

$$K(n,p) = \begin{cases} \displaystyle\sum_{\hat{n}\in\mathscr{N}} g_d(n,\hat{n})\, e^{-j\hat{n}(p-q_c)\frac{2\pi}{N}} & \text{for } p \in \mathscr{P}_D \\ \displaystyle\sum_{\hat{n}\in\mathscr{N}} g_n(n,\hat{n})\, e^{-j\hat{n}(p+q_c)\frac{2\pi}{N}} & \text{for } p \in \mathscr{P}_N \\ K_v(n,p) & \text{for } p \in \mathscr{P}_V \end{cases} \tag{4.33b}$$

with $n \in \mathscr{N}$, $\mathscr{P}_D = \{0,\ldots,2q_c-1\}$, $\mathscr{P}_N = \{P-2q_c,\ldots,P-1\}$ and $\mathscr{P}_V = \{2q_c,\ldots,P-2q_c-1\}$. Defining $W_N^{np} = \frac{1}{N}e^{jnp\frac{2\pi}{N}}$, and the matrices

$$\left[\hat{k}_d\right]_{pn} = W_N^{n(p-q_c)},\ p \in \mathscr{P}_D,\ \left[\hat{k}_v\right]_{pn} = k_v(p,n),\ p \in \mathscr{P}_V,$$
$$\left[\hat{k}_n\right]_{pn} = W_N^{n(p+q_c)},\ p \in \mathscr{P}_N,\ \left[H_{[\boldsymbol{\theta}]_i}\right]_{n\hat{n}} = h_{[\boldsymbol{\theta}]_i}(n,\hat{n})$$

where $(n,\hat{n}) \in \mathscr{N}\times\mathscr{N}$, and H_d, $H_n \in \mathbb{R}^{N\times N}$, then k, defined in section §3.2.2.2, can be partitioned as:[6]

$$k^T = \left[\, H_d\,\hat{k}_d^T \;\middle|\; H_n\,\hat{k}_n^T \;\middle|\; \hat{k}_v^T \,\right] \tag{4.34}$$

[6]Note $\hat{k}_d$ and $\hat{k}_n$ are identical; their explicit form is shown to emphasis that for separability, $H_{[\boldsymbol{\theta}]_i}\hat{k}_{[\boldsymbol{\theta}]_i}$ span different regions of the p-domain.

where $\hat{k}_v$ are the unused basis vectors. Since the Fourier kernel is an orthonormal basis, then $\{\hat{k}_{[\boldsymbol{\theta}]_i}\}$ has full rank and:

$$\text{rank}\,[\hat{k}_d] = \text{rank}\,[\hat{k}_n] = 2q_c \text{ and } \text{rank}\,[\hat{k}_v] = P - 4q_c$$

As discussed in section §4.4.2, if separation is possible, the kernels must be transform pairs and, therefore, k must have full rank. Ergo, the problem now is to find constraints on the matrices H_d and H_n such that k satisfies this condition.

Theorem 21 (Discrete-Time Separability Criterion) *The signals* $\mathbf{d}(n)$ *and* $\mathbf{n}(n)$ *are separable if the known, and completely specified, matrix* k *of* (4.34) *is of full rank.*

One physical way of interpreting the rôle of H_d and H_n in equation (4.34) is to consider H_d and H_n as linear transformations of $\mathbf{a}(n)$ and $\mathbf{b}(n)$, mapping them to *different subspaces*. Thus, $\mathbf{d}(n)$ and $\mathbf{n}(n)$ are not separable when these linear transformations map $\mathbf{a}(n)$ and $\mathbf{b}(n)$ to *overlapping subspaces*. Clearly, if $H_d = H_n$, separation is not possible because no addition signal structure has been specified – both signals are mapped to the same subspace. Using any additional information regarding the signal structure allows a transformation to be designed which maps $\mathbf{a}(n)$ and $\mathbf{b}(n)$ to disjoint subspaces.

4.5.3 Minimum Sampling Frequency

It can be deduced from section §4.5.2 that in order to separate two modulated signals, the kernel must span a subspace with rank of at least $4q_c$. Therefore, there must be at least $4q_c$ data samples and, hence, it follows that the separation of stochastic signals requires them to be oversampled by a factor of at least 2.

4.6 Uniform Modulation

Separation of the special class of nonstationary signals known as *uniformly modulated processes*, or amplitude modulated time series, characterised by the form

$$\mathbf{y}(t) = c(t)\,\mathbf{x}(t), \qquad t \in \mathscr{T} \tag{4.35}$$

where $\mathbf{x}(t)$ is a stochastic process and $c(t)$ is a known deterministic signal [296], is of great importance and is studied in this section using the method employed in

section §4.5. Such nonstationary processes have been shown to describe the behaviour of seismic reflectivity data [115, 296] where, for example, earthquake and explosion data have been modelled by equation (4.35), with $\mathbf{x}(t)$ as a zero mean AR(1) process, and the deterministic function given by $c(t) = t^{\alpha}\nu^{-\beta t}$, with α, β and ν as known constants. The model in equation (4.35) is also appropriate for modelling deterministic signals that have been corrupted by multiplicative noise, for example, in radar applications where a target is illuminated and the signal experiences amplitude distortion caused either by target fluctuation, or scattering of the medium (*e.g.* fading) [36, 37]. This manifests itself as a random time-varying amplitude, $c(t)$, which can be viewed as an unwanted phenomenon; the terminology *multiplicative noise* is often used in the literature. Ghogho and Garel [115] investigate the estimation of the deterministic process, $c(t)$, for a time series when $\mathbf{x}(t)$ is a stationary autoregressive process. Other examples of uniformly modulated processes may be found in quadrature and spread spectrum modulation schemes, with the former considered in section §4.6.1.

Consider the special case of (2.1a) when

$$\left.\begin{aligned} h_d(t,\tau) &\equiv h_d(t)\,\delta(t-\tau) \\ h_n(t,\tau) &\equiv h_n(t)\,\delta(t-\tau) \end{aligned}\right\} \quad t \in T \tag{4.36a}$$

such that $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are modelled as uniformly modulated processes:

$$\left.\begin{aligned} \mathbf{d}(t) &= h_d(t)\,\mathbf{a}(t) \\ \mathbf{n}(t) &= h_n(t)\,\mathbf{b}(t) \end{aligned}\right\} \quad t \in T \tag{4.36b}$$

where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are stochastic processes bandlimited to $\pm\omega_c$. The first two expressions of the transform kernel in (4.31a) reduce to the form:

$$k(\lambda,t) = \begin{cases} \frac{1}{2\pi}\,h_d(t)\,e^{j(\lambda-\omega_c)t} & \text{for } \lambda \in [0, 2\omega_c) \\ \frac{1}{2\pi}\,h_n(t)\,e^{j(\lambda+\omega_c)t} & \text{for } \lambda \in [-2\omega_c, 0) \end{cases} \tag{4.37}$$

The form of the function $k_\nu(\lambda,t)$ in (4.31a) can be deduced by writing $h_d(t)$ and $h_n(t)$ as Fourier transform:

$$h_d(t) \triangleq \frac{1}{2\pi}\int_{\hat{\omega}\in\Omega} H_d(\hat{\omega})\,e^{j\hat{\omega}t}\,d\hat{\omega} \quad \text{and} \quad h_n(t) \triangleq \frac{1}{2\pi}\int_{\hat{\omega}\in\Omega} H_n(\hat{\omega})\,e^{j\hat{\omega}t}\,d\hat{\omega} \tag{4.38}$$

such that equation (4.37) becomes:

$$k(\lambda, t) = \begin{cases} \dfrac{1}{(2\pi)^2} \displaystyle\int_{\hat{\omega}\in\Omega} H_d(\hat{\omega})\, e^{j(\lambda-\omega_c+\hat{\omega})t}\, d\hat{\omega} & \text{for } \lambda \in [0, 2\omega_c) \\ \dfrac{1}{(2\pi)^2} \displaystyle\int_{\hat{\omega}\in\Omega} H_n(\hat{\omega})\, e^{j(\lambda+\omega_c+\hat{\omega})t}\, d\hat{\omega} & \text{for } \lambda \in [-2\omega_c, 0) \end{cases} \tag{4.39}$$

which is a linear combination of the basis set $\{e^{j\omega t},\ \forall\omega \in \hat{\Omega} \subset \Omega\}$. To complete the basis set, $k_\nu(\lambda, t)$, which is itself a linear combination of the set $\{e^{j\omega t},\ \forall\omega \in \Omega\}$, must not be a linear combination of $k(\lambda, t)$ for $\lambda \in (-2\omega_c, 2\omega_c)$. It can thence be deduced that if $k_\nu(\lambda, t) = e^{jg(\lambda)t}$, for some function $g(\lambda)$, the basis set can be completed provided that $k(\lambda, t)$ is not degenerate in the range $\lambda \in (-2\omega_c, 2\omega_c)$. In many cases, $\{k_\nu(\lambda, t) = e^{2j\lambda t},\ \forall\lambda \notin (-2\omega_c, 2\omega_c)\}$ suitably extends the basis. In such a case, the expression for the kernel $k(\lambda, t)$ reduces to:

$$k(\lambda, t) = \begin{cases} \frac{1}{2\pi} h_d(t)\, e^{j(\lambda-\omega_c)t} & \text{for } \lambda \in [0, 2\omega_c) \\ \frac{1}{2\pi} h_n(t)\, e^{j(\lambda+\omega_c)t} & \text{for } \lambda \in [-2\omega_c, 0) \\ \frac{1}{2\pi} e^{2j\lambda t} & \text{for } \lambda \notin (-2\omega_c, 2\omega_c) \end{cases} \tag{4.40}$$

Moreover, assuming that the set of exponential basis functions for the unused transform space completes the set of basis functions then, since the rank of the kernel matrix can readily be determined, the inverse kernel matrix in (4.34) can be readily constructed and its inverse calculated, facilitating the separation of separable discrete–time signals. This approach is outlined in section §4.6.2 for amplitude modulated chirp signals. To motivate the assumption that the exponential set can be used to complete the basis set, the kernel arising in quadrature modulation is discussed in the following example.

4.6.1 Separation of Quadrature Modulated Signals

In Example 6 on page 84 it was shown that the two quadrature modulated signals

$$\left.\begin{aligned} \mathbf{d}(t) &= \sin\omega_c t\, \mathbf{a}(t) \\ \mathbf{n}(t) &= \cos\omega_c t\, \mathbf{b}(t) \end{aligned}\right\} \quad t \in \mathscr{T} \tag{4.9}$$

where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are bandlimited stochastic signals with Fourier spectral components in the range $(-\omega_c,\ \omega_c)$, can be separated using the quadrature demodulator

in Figure 4.1 on page 84, which is an ideal filter with LTV impulse response:

$$h(t,\tau) = 2\frac{\sin\omega_c\,(t-\tau)}{\omega_c\,(t-\tau)}\sin\omega_c t\ \sin\omega_c\tau \tag{4.11}$$

The derivation of this filter may also be obtained from the concatenated transform method. The modulating filters in equation (2.1a) become:[7]

$$\left.\begin{aligned} h_d(t,\tau) &= \sin\omega_c t\,\delta(t-\tau)\\ h_n(t,\tau) &= \cos\omega_c t\,\delta(t-\tau)\end{aligned}\right\}\quad (t,\tau)\in\mathbf{T} \tag{4.41}$$

Furthermore, it can be shown that the transform kernel $k(\lambda,t)$ is given by

$$k(\lambda,t) = \begin{cases} \frac{1}{2\pi}\sin\omega_c t\,e^{j(\lambda-\omega_c)t} & \text{for } \lambda\in[0,2\omega_c)\\ \frac{1}{2\pi}\cos\omega_c t\,e^{j(\lambda+\omega_c)t} & \text{for } \lambda\in[-2\omega_c,0)\\ \frac{1}{2\pi}e^{2j\lambda t} & \text{for } \lambda\notin(-2\omega_c,2\omega_c)\end{cases} \tag{4.42}$$

and $K(t,\lambda)\propto k^*(\lambda,t)$, which can be verified by substitution into (3.10). To recover $\mathbf{d}(t)$, the pass-band will be $[0,2\omega_c)$, and the ideal filter is given by:

$$h(t,\tau) = \int_0^{2\omega_c} k(\lambda,t)\,K(\tau,\lambda)\,d\lambda \tag{4.8}$$

which also gives the perfect filter in equation (4.11).

4.6.2 Separation of Chirp Modulated Signals

Chirp signals embedded in multiplicative noise is a topic of considerable interest in many practical situations, such as radar and speech. Consider, again, the radar application where a target is illuminated; due to the relative motion between the target and the receiver, the phase of the transmitted sinusoidal signal will be shifted, and this phase shift can be adequately modelled as

$$\phi(t) = a_0 + a_1 t + a_2 t^2 \tag{4.43}$$

provided that the motion is continuous and differentiable [37]. The transmitted sinusoid is reflected as a chirp signal, and this will be distorted by multiplicative noise as discussed in section §4.6.

[7] Compare the expressions for the modulating filters in (4.41) with the response in (4.10).

Now consider the speech process: during the generation of phonemes such as vowels, shifts in the 'centre' frequency of a formant occurs, as well as changes in their bandwidth (see section §2.2.3). A simplified model of this nonstationary process, when considering just a single formant, could be a narrow-band limited nonstationary stochastic process modulated by a chirp signal, where the frequencies swept out by the chirp are in the region of bandwidth of the modulated signals. The modelling of speech using frequency-varying sinusoids, with a chirp as a special case, has been considered by Marques and Almeida [244] and this model is also apparent in [58]. It is therefore of interest of consider the separation of chirp signals distorted by multiplicative noise, and this is studied in the following sections.

4.6.2.1 Problem Formulation

Suppose that the desired and noise signals can be expressed in the form

$$\left.\begin{aligned}\mathbf{d}(t) &= h_d(t)\,\mathbf{a}(t)\\ \mathbf{n}(t) &= h_n(t)\,\mathbf{b}(t)\end{aligned}\right\} \quad t \in \mathscr{T} \tag{4.36b}$$

where $\mathscr{T} = \{0, \ldots, T-1\}$, and the deterministic functions, $h_d(t)$ and $h_n(t)$, are given by:

$$\left.\begin{aligned}h_d(t) &= \cos\left[2\pi p_i \frac{t}{f_s}\left(1 + \frac{1}{\tau_d}\frac{p_f - p_i}{p_i}\frac{t}{f_s}\right)\right]\\ h_n(t) &= \cos\left[2\pi q_i \frac{t}{f_s}\left(1 + \frac{1}{\tau_n}\frac{q_f - q_i}{q_i}\frac{t}{f_s}\right)\right]\end{aligned}\right\} \quad t \in \mathscr{T} \tag{4.44}$$

where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are bandlimited to q_c, the sampling frequency $f_s \geq 4q_c$, the *start* and *stop* sweep frequencies are given by $(p_i,\ q_i)$ and $(p_f,\ q_f)$ respectively, and the frequency sweep times are $(\tau_d,\ \tau_n)$. The arbitrary phase term in each of the cosine functions, $h_d(t)$ and $h_n(t)$, are set to zero for clarity.

4.6.2.2 Separability Constraints

If $\{h_d(t) = h_n(t), \forall t \in \mathscr{T}\}$, separation is not possible since no additional signal structure distinguishing $\mathbf{d}(t)$ and $\mathbf{n}(t)$ is specified. However, by a continuity argument, it is reasonable to assume that a small perturbation in one of the functions is not likely to make separation possible either. Furthermore, if $q_c \gg (p_i,\ q_i,\ p_f,\ q_f)$, such that $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are only slightly modulated, it is also reasonable to assume

that there are not enough distinguishing features to achieve separation. These intuitive continuity arguments suggest that the signals $\mathbf{d}(t)$ and $\mathbf{n}(t)$ can be separated from their mixture provided:

1. The spectra of the modulating signals, $h_d(t)$ and $h_n(t)$, differ *sufficiently*, but are not, in general, disjoint.

2. The boundary frequencies, (p_i, q_i) and (p_f, q_f), are in the *region* of q_c: *i.e.* to ensure $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are modulated such that $\mathbf{d}(t)$ and $\mathbf{n}(t)$ have different significant characteristics, albeit, not necessarily in the Fourier domain.

4.6.2.3 Results

Figure 4.3(a) shows the Fourier spectra of the modulating signals $h_d(t)$ and $h_n(t)$ of (4.44) with $p_i = q_f = 4250$ Hz, $p_f = q_i = 5750$ Hz, $f_s = 22.1$ kHz, and $\tau_d = \tau_n = 70$ ms: these sweeps are identical, except one is in the opposite direction to the other. Consider taking $N = 1000$ samples or 45 ms worth of data, with $q_c = 5083$ Hz (230 frequency bins), such that there will be 80 unused generalised frequency bins. The kernel matrix

$$k^T = \left[H_d \hat{k}_d^T \;\middle|\; H_n \hat{k}_n^T \;\middle|\; \hat{k}_v^T \right] \tag{4.34}$$

with $H_d = \operatorname{diag}[h_d(t)]$, $H_n = \operatorname{diag}[h_n(t)]$ and $[k_v]_{pt} = W_N^{tp}$, $p \in \mathscr{P}_V$, $t \in \mathscr{T}$, has full rank and, therefore, the signals are separable. The inverse matrix $K = k^{-1}$ is easily calculated and, since the *acceptance* region is $\mathscr{P}_D = \{0, \ldots, 2q_c - 1\}$, equation (4.18a) can be used to calculate the impulse response of the ideal filter. In this case, the resulting filter response, $h(t, \hat{t})$, is shown in Figure 4.4(d). It is interesting to note that the response of the filter is, in fact, complex, although when the input to the filter is real, the imaginary part of the filter can be ignored: Figure 4.4(d) actually only shows the real part of the filter's response. Notice that the filter is operating on a finite temporal domain, therefore introducing some artifacts in the impulse response, particularly around the 'edges' of the plot: the response is clearly time-varying.

4.6.2.4 Analysis

Since the Fourier spectra of the modulating signals in Figure 4.3(a) overlap, so do the Fourier spectra of $\mathbf{d}(t)$ and $\mathbf{n}(t)$. To emphasis this, consider the unknown ban-

dlimited signals $\mathbf{a}(t)$ and $\mathbf{b}(t)$ shown in Figure 4.3(b), with their Fourier spectra shown in Figure 4.3(c). The resulting desired and noise signals, $\mathbf{d}(t)$ and $\mathbf{n}(t)$, are shown in Figure 4.3(d), with their Fourier spectra shown in Figure 4.4(a). These spectra are clearly overlapping, and the Fourier spectrum of $\mathbf{x}(t)$ is shown in Figure 4.4(b). Therefore, conventionally, $\mathbf{d}(t)$ and $\mathbf{n}(t)$ would be considered inseparable; however, taking a generalised spectral transform using equation (3.16) with the kernel matrices derived in equation (4.34), gives the generalised spectrum of $\mathbf{x}(t)$ shown in Figure 4.4(b). This generalised spectrum can then be bandpass filtered to recover the generalised spectrum of either $\mathbf{d}(t)$ or $\mathbf{n}(t)$, and the inverse transform can be taken to obtain the time–domain representations of the signals. Since the kernel matrix has full rank, *perfect separation* has been achieved. Note, from section §4.5.2, that the generalised spectra of the signals $\mathbf{d}(t)$ and $\mathbf{n}(t)$ have values corresponding to those in the shifted Fourier frequency spectra. In other words, $\mathbf{D}(p) = \mathbf{A}(p - q_c)$ and $\mathbf{N}(p) = \mathbf{B}(p + q_c)$: ergo the resemblance between the generalised spectra of Figure 4.4(c) and that of Figure 4.4(b), although the Fourier and generalised spectral domains are, of course, entirely different. Moreover, note that the generalised frequency bins 980 to 1000 are *empty* since there are no signal components in this region of the generalised frequency domain.

4.6.2.5 Required Prior Knowledge

The *prior* knowledge required to separate the signals is:

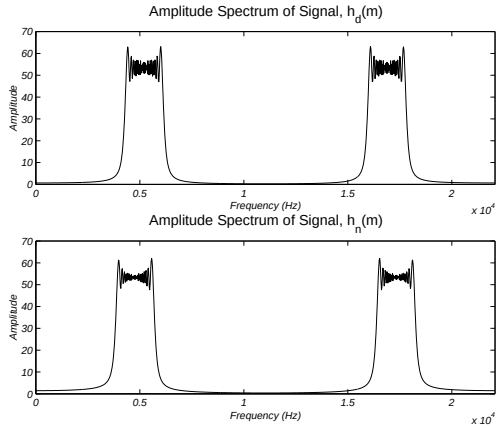
1. The signals $\mathbf{d}(t)$ and $\mathbf{n}(t)$ must be known to be of the form (4.36b) with $h_d(t)$ and $h_n(t)$ given by (4.44).
2. The signals $\mathbf{a}(t)$ and $\mathbf{b}(t)$ must be known to be bandlimited to q_c.
3. The sampling frequency $f_s \geq 4q_c$, *start* and *stop* sweep frequencies denoted by (p_i, q_i) and (p_f, q_f), respectively, and the frequency sweep times (τ_d, τ_n) in equation (4.44) must all be known.

These parameters must somehow be estimated from the observed signal and, although this is a non-trivial problem, for the purpose of signal separation it is reasonable to assume that these few parameters are known *a priori*. The problem of estimating these chirp parameters for chirp signals in white noise has been considered by, for example, Djurić *et al*. [93,214], and similar techniques could be developed to estimate the parameters if they are unknown; as previously mentioned, this task is left for future work as it is dependent on the particular problem at hand.

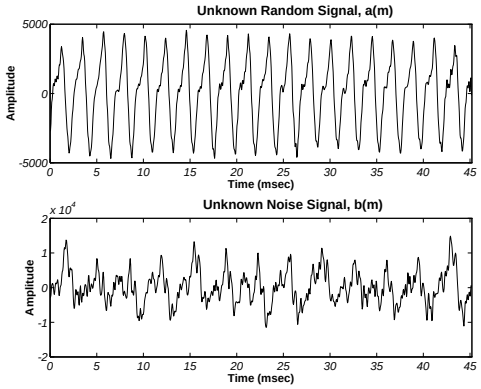
4.7 Chapter Summary

By representing a signal on another domain, separability criteria become apparent from this new viewpoint. For example, it is well know that stationary stochastic signals are separable if, and only if, they have non-overlapping Fourier spectra. However, from a time-domain perspective, it is difficult to derive separability conditions. The concept of an ideal filter has been introduced to separate the sum of two signal classes, leading to separability conditions for *abstract* classes of signals. By definition, the ideal filter passes, without distortion, all signals belonging to an acceptance class, and rejects all other signals which belong to a rejection class. Such a filter can be implemented by reformulating this definition in terms of spectral components. Hence, as commonly defined, an ideal filter is one which passes, without distortion, all spectral components within a generalised frequency range, but does not pass spectral components outside this range. Since this definition is in terms of spectral components, it leads to an expression for the structure of an ideal filter in terms of the basis functions defining the spectral domain on which the filter is ideal. It has also been demonstrated that a sufficient condition for obtaining perfect separation using the nonstationary Wiener filter is that it must be an ideal filter. Implicitly, therefore, the Wiener filter equations and its solution have linked the method of representing a stochastic process on an arbitrary domain as a power spectrum, to that of representing a signal as a stochastic spectral process. Furthermore, this link has highlighted the *stochastic* nature of using an ideal filter to separate stochastic signals.

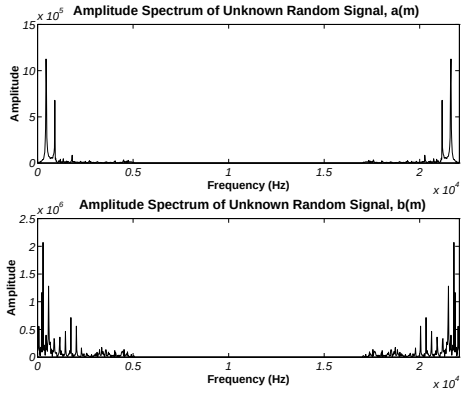
This chapter has also introduced a signal separation technique by *concatenating* the domains on which two signal classes can be represented on with a finite number of basis functions. These signal domains should, in practice, be inferred by some *limited* prior knowledge regarding the structure of the signal. The technique has been specifically applied to uniformly modulated signals, and an example has been discussed of separating chirp signals embedded in multiplicative noise. This has applications in many practical situations, such as radar and speech. Filtered and uniformly modulated signals that overlap in the Fourier domain are often assumed to be inseparable; if the modulating (bandlimited) functions $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are unknown, but the modulated functions $\mathbf{h}_{[\boldsymbol{\theta}_i]}(t)$ are known, then provided (4.34) is satisfied, the resulting processes can, in fact, be separated.



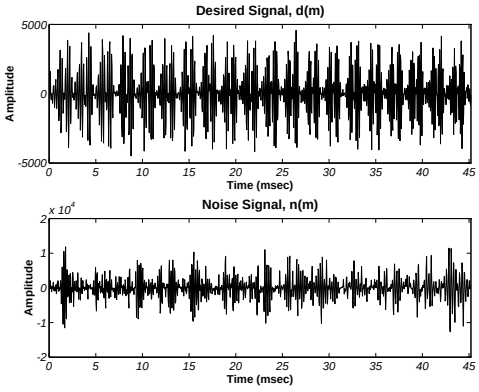
(a) Fourier spectra of $h_d(t)$, $h_n(t)$; $p_i = q_f = 4250$ Hz, $p_f = q_i = 5750$ Hz, $f_s = 22.1$ kHz, $\tau_d = \tau_n = 70$ ms.



(b) Unknown stochastic signals $\mathbf{a}(t)$, $\mathbf{b}(t)$, sampled at 22.1 kHz and bandlimited to 5.083 kHz.

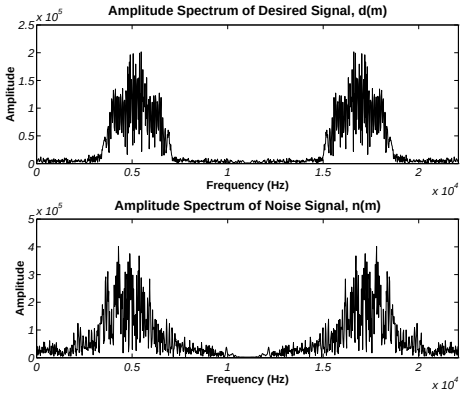


(c) The Fourier transforms of the time series in Figure 4.3(b), emphasising their bandlimited nature.

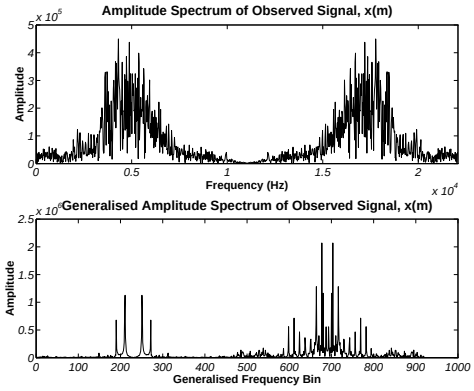


(d) The desired and noise signals resulting from multiplying the signals in Fig. 4.3(b) by the chirp signals of Fig. 4.3(a).

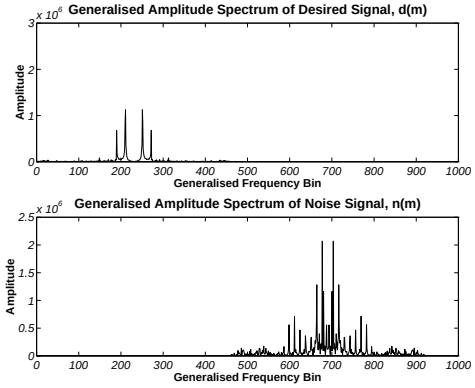
Figure 4.3: *The form of the signals used in the 'chirp' example of signal separation.*



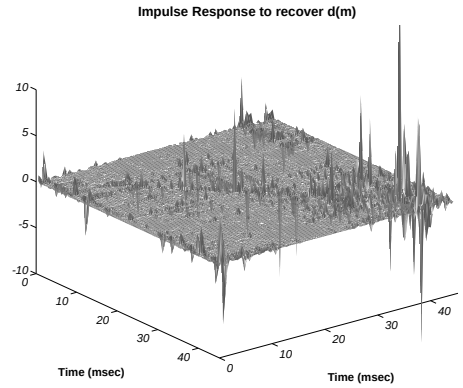
(a) The Fourier transform of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ in Figure 4.3(d); the spectra of $\mathbf{d}(t)$ and $\mathbf{n}(t)$ are clearly overlapping, and so conventionally, these would be considered inseparable.



(b) *Top:* The Fourier transform of the observed signal $\mathbf{x}(t) = \mathbf{d}(t) + \mathbf{n}(t),\ \forall t \in \mathcal{T}$ for the resulting signals in Figure 4.3(d); *Bottom:* The Generalised transform of $\mathbf{x}(t)$.



(c) The Generalised Spectra of $\mathbf{d}(t)$ and $\mathbf{n}(t)$, obtained by bandpass filtering the Generalised Spectrum of $\mathbf{x}(t)$ in Figure 4.4(b) with pass-bands $\{0, 2q_c - 1\}$ and $\{2q_c, 4q_c - 1\}$ respectively.



(d) The response of the ideal filter which recovers $\mathbf{d}(t)$ for the chirp modulated case discussed in section §4.6.2.1.

Figure 4.4: *The Fourier and Generalised Spectra of signals used in the 'chirp' signal separation example, and the form of the impulse response of the* ideal filter.

5

Part II Conclusions: Signal Separation

NONSTATIONARITY is useful in applications where the mixture of two or more signals is observed and, at each time instance, there is only a *single* observation of the mixture in the time domain. Separation of the signals has important applications where background noises must be removed in, for example, mobile telephones, forensic science, restoration of black box recordings, speech recoginition, and biomedical data analysis. In particular, improving the intelligibility of speech in hearing aids and conference environments is complicated by the presence of unwanted disturbances, or distortions of speech, such as background noise, ranging from environmental noise to other conversations, and other signal degradations which interfere with the desired sound. These disturbances must be suppressed or reduced to adequately low levels.

Signal separation with one observation sensor can only be achieved by exploiting prior knowledge of the signal structure, since the separation problem is inherently under-constrained. Moreover, it is desirable that this knowledge be specified such that it is common to a 'class' of stochastic process. For example, it is well

known that stationary stochastic signals are 'perfectly' separable if their power spectra do not overlap in the Fourier domain. Separation can be achieved using a bandpass filter when the pass-band frequencies are known *a priori*. It is not, however, necessary to know the actual values of the spectral components of the two signals. Similarly, nonstationary stochastic signals that are non-overlapping in the time domain can be separated using a temporal switch, which is a special case of a LTV filter, provided the times for which they have non-zero components are known. Generalising, it follows that if an arbitrary signal domain exists such that the representation of two classes of signals are disjoint, then signal separation can be achieved using a generalised LTV *bandpass filter*, first proposed by Zadeh [413–415, 418].

The separability of signal mixtures has been defined in section §3.1.1 as the identification of the accuracy to which signals can be separated when, at each time instance, only one observation of the mixture in the time domain is available. Using the assumption that signal separation is possible using linear time-varying filters, or the generalised Wiener-Hopf filter, an intuitive solution to the Wiener filter equations is obtained in section §4.2, and relies on the factorisation of the autocorrelation functions into *generalised power spectra*. Chapter 3 defines the generalised power spectrum, and is based on the 'stochastic spectral representation' of a stochastic process. The generalisation of the power spectrum for nonstationary processes leads to a generalised signal domain, and two signals can be 'perfectly' separated if such a domain can be found such that a desired signal and a noise signal have components that lie in disjoint regions. The solution to the Wiener-Hopf filter equations leads naturally to a criterion for the separation of nonstationary signals with perfect precision, and is a natural extension to the condition for separating stationary signals. This filter is composed of a term independent of the signal values, corresponding to regions in the spectral domain where the desired signal components are not distorted by interfering noise components, and a term dependent on the signal correlations, corresponding to the region where components overlap.

Chapter 4 has also introduced a signal separation technique by *concatenating* the domains on which two signal classes can be represented using a finite number of basis functions. These signal domains should, in practice, be inferred by some *limited* prior knowledge regarding the structure of the signal. The technique has been specifically applied to uniformly modulated signals, and an example has been discussed of separating chirp signals embedded in multiplicative noise. This has applications in many practical situations, such as radar and speech.

III

Single Channel Blind Deconvolution of Random Signals with Infinite Support

6

Dereverberation of Acoustic Environments

IN general, acoustic signals radiated within a room are linearly distorted by reflections off walls and other objects [8, 59, 111, 143, 168, 191, 201, 262, 354]. These distortions, which arise as a result of this *reverberation effect*, often spoil speech intelligibility in devices such as 'hands-free' conference telephones, automatic speech recognition, and hearing aids. For example, as discussed in Chapter 1, reverberation and *spectral coloration* cause users of hearing aids to complain of being unable to distinguish one voice from another in a crowded room [99, 100, 289, 317]. It is therefore of significant importance to investigate the application of signal processing techniques for the enhancement of the quality of speech distorted in an acoustic environment. The techniques employed to achieve this speech enhancement consist of two stages: first, estimation of the properties of the acoustic environment from the observed reverberant speech and, second, given both the reverberant speech and the acoustic properties of the room, estimation of the original speech. The first stage of this process is discussed in depth in Chapters 7 to 9 in Part III of this dissertation. In order to achieve *blind dereverberation* in

an acoustic environment, it must be demonstrated that *dereverberation* is possible when the properties of the acoustic environment are known *a priori*. The purpose of this introductory chapter is to investigate existing dereverberation techniques to prove the validity of the proposed approach to blind speech dereverberation.

6.1 Room Acoustics

This section introduces some basic theoretical properties of acoustics which are important for understanding why particular models are used throughout this work. The characteristic properties of a particular acoustic environment depend on the frequency components of the sound of interest; the various techniques which are used for the analysis of these properties are discussed in section §6.1.1. The room transfer function (RTF) is introduced, and reverberation time (RT) discussed as a measurement of the severity of reverberation within a room.

6.1.1 Analysing Room Acoustics

There are many different techniques for analysing the acoustics of a room and, in general, each of these techniques applies to a different frequency range of the audible spectrum; no single analytic or numerical tool can currently model the entire audible frequency range between 16 Hz and 15 kHz [112, 201]. The audible spectrum can be divided into four regions, over each of which a different analytical tool is appropriate.

6.1.1.1 Very Low Frequencies

If the frequency of a sound source is below $f_w = \frac{c}{2L}$, where c is the speed of sound, and L is the largest dimension of the acoustic environment, there is no resonant support for the sound in the room. This frequency band can be analysed using non-harmonic solutions to the *wave equation* (see equation (6.2) in section §6.1.2).

6.1.1.2 Comparable Room Dimensions and Wavelength: Wave Acoustics

The next region corresponds to frequencies for which the wavelength of the sound in consideration is comparable to the dimensions of the room. This region spans

from the lowest resonant mode to the *Schroeder cut-off frequency* [191,201]:

$$f_g \approx \frac{5000}{V\bar{\delta}} \quad (\text{Hz}) \tag{6.1a}$$

where V is the volume of the room in m^3, and $\bar{\delta}$ is the *mean* value of the damping constants associated with each resonant in the room (see equation (6.3) in section §6.1.2). The mean damping constant is related to the *reverberation time*, discussed in section §6.1.3.3, by the relation $T_{60} = 6.91/\bar{\delta}$ (see equation (6.6)), such that the cut-off frequency can be expressed in the well known form:

$$f_g \approx 2000\sqrt{\frac{T_{60}}{V}} \quad (\text{Hz}) \tag{6.1b}$$

This is the lower frequency limit at which a statistical treatment of superimposed vibrational modes in a room is permissible. In this region, *wave acoustics* are applicable for describing the acoustical properties of a room. Wave acoustics assume a harmonic sound source and are based on solutions of the *wave equation* of equation (6.2). For instance, in an small living room with dimensions $3 \times 5 \times 7\,\text{m}$ and a reverberation time of 0.5 sec, statistical theory would be relevant above 138 Hz.

6.1.1.3 Very High Sound Frequencies: Geometrical Room Acoustics

At very high sound frequencies *geometrical room acoustics* apply. As in geometrical optics, geometrical room acoustics employs the limiting case of vanishingly small wavelengths. This assumption is valid if the dimensions of the room and its walls are large compared with the wavelength of the sound; a condition which is met for a wide-range of audio frequencies in standard rooms. Hence, in this frequency range, specular reflections and the *sound ray*[1] approach to acoustics prevail. Geometrical acoustics usually neglect wave related effects such as diffraction and interference.

6.1.1.4 High Sound Frequencies

The final region, or *transition region*, consists of the frequency components between f_g and, approximately, $4f_g$, where f_g is given by equation (6.1). In this region, the wavelengths are often too short for accurate modelling using *wave acous-*

[1] A sound ray is meant as a small portion of a spherical wave with vanishing aperture, which originates from a certain point. It has a well-defined direction of propagation and is subject to the same laws of propagation as *light rays*, apart from the different propagation attenuation [191,201].

tics, and too long for *geometric acoustics*. Thus, in general, a statistical treatment is employed. The boundary frequencies of this band for typical acoustic environments can be calculated using equation (6.1); for example, a small recording studio of dimensions $3 \times 5 \times 7$ m and $T_{60} = 0.5$ sec gives a transition region of 138 Hz to 552 Hz, whilst a car compartment of volume $V = 2.5\,m^3$ and $T_{60} = 0.05$ sec gives a region of 282 Hz to 1131 Hz [112].

6.1.2 Room Transfer Function

In principle, any complex sound field can be considered as a superposition of numerous simple sound waves, *e.g.* plane waves [191, 201], and their propagation within a room can be considered linear if the medium in which the waves travel is assumed to be homogeneous, at rest, and independent of wave amplitude. Under these assumptions, when the acoustical power of the sound source is doubled, the sound pressure produced by the sound source at some distant point doubles in value; this property has its mathematical analogue in the linearity of the wave equation [191, 201]:

$$\begin{aligned} \frac{\partial^2 p}{\partial t^2} &= c^2 \nabla^2 p \\ &= c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right) \end{aligned} \tag{6.2}$$

In practice, the effects of temperature variations cause the velocity of sound to be a function of time and spatial position. Furthermore, the air is not completely at rest due to temperature differences and air conditioning and, therefore, solutions of the wave equation (6.2) will contain nonlinearities. However, the effects of these inhomogeneities are so small that they can be ignored.

It can be shown, using the solutions of the wave equation (6.2) for sound fields in a *closed space*, that the transfer function between a sound source and receiver, with spatial coordinates denoted by position vectors $\mathbf{r}_s$ and $\mathbf{r}_o$ respectively, can be expressed in terms of the resonant frequencies, ω_i, and their eigenfunctions, $P_i(\mathbf{r})$,[2]

[2]The eigenfunctions are mutually orthogonal, and satisfy:

$$c \iiint_V P_n(\mathbf{r})\, P_m(\mathbf{r})\, dv = \begin{cases} \omega_n + j\,\delta_n & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

where the integration is performed over the volume V enclosed by the walls. At angular frequency $\omega = \omega_n$, the associated term in (6.3) assumes a particularly high absolute value and, as such, the

as [201]:

$$H_{(\mathbf{r}_s,\mathbf{r}_o)}(\omega) = G \sum_{i=1}^{\infty} \frac{P_i(\mathbf{r}_s)\, P_i(\mathbf{r}_o)\, j\omega}{\omega^2 - \omega_i^2 + \delta_i^2 - 2j\delta_i\omega_i} \tag{6.3}$$

where ω is the angular frequency, δ_i is the damping constant (corresponding to the Q-factor), and G is a gain constant. The parameters ω_i and δ_i are independent of the source and receiver positions, and their values are determined by the room size, wall reflection coefficient, and room shape. The frequency response of the room described by equation (6.3) leads to an acoustic impulse response (AIR) or, alternatively, room impulse response, $h_{(\mathbf{r}_s,\mathbf{r}_o)}(t)$. The variation of the acoustic impulse response, or room transfer function, with source and observer positions $(\mathbf{r}_s, \mathbf{r}_o)$ is discussed in [80, 257], and its variation with temperature in [276]. The form of equation (6.3) leads to the justification for the use of some well know modelling techniques used in signal processing, as discussed later in this chapter.

6.1.3 Reverberation

Reverberation is a phenomenon which plays a major rôle in every aspect of room acoustics, and which yields the least controversial criterion for the judgement of the acoustical qualities of a room, or an equalised recording made in a reverberant environment [201]. Throughout this work, reverberation will be considered as the sum total of all sound reflections arriving at a certain point in a room after the room has been excited by an impulsive sound signal. Reverberation can also be considered as the common decaying of free vibrational modes within a room when considered from a statistical viewpoint [191, 201, 230, 249, 250]. Although the latter approach may seem more applicable for a treatise on statistical signal processing, it is, in fact, more appropriate to consider the former method which has its mathematical analogue in the finite impulse response (FIR) and infinite impulse response (IIR) representations.

6.1.3.1 Early and Late Reflections

The room transfer function of (6.3) yields a corresponding *room impulse response*. The reverberant component of this impulse response can be divided into two components; *early reflections* and *late reflections*. Early reflections are not perceived

corresponding frequencies, ω_n, are called *eigenfrequencies* of the room, and is sometimes referred to as *resonant frequencies* due to the sort of resonances occurring in the vicinity of the ω_n's.

as a separate sound to the direct sound so long as the delay does not exceed a certain limit – a perceptual characteristic often referred to as the *precedence effect* [72, 226]; although the direct sound is followed by multiple reflections, which would be audible in isolation, the first-arriving wavefront dominates many aspects of perception independent of the energy of these early reflections. As such, these early reflections are actually perceived to reinforce the direct sound and are therefore considered useful with regards to speech intelligibility.[3] Reflections which arrive with larger delays w. r. t. the arrival of the direct sound are perceived either as separate echoes, or as reverberation; as such *late reflections* impair speech intelligibility [387]. It should also be noted that, from a signal processing perspective, *early reflections* appear as separate delayed and attenuated impulses in the room impulse response, whilst *late reflections* appear as a continuum.

6.1.3.2 Performance Criteria

There are a number of time-domain criteria for indicating the intelligibility of filtered speech for a given impulse response. Bradley [51] has evaluated a range of acoustical measures as predictors of speech intelligibility scores, and two are considered here. The first criterion assumes that the known sampled impulse response $\{h(t),\ t \in \mathscr{T} \subset \mathbb{Z}^+\}$ consists of a component, $d(t)$, corresponding to the direct path of sound travel,[4] and a reverberant signal component, $g(t)$, such that:

$$h(t) = d(t) + g(t)\,, \quad \forall t \in \mathscr{T} \tag{6.4}$$

If the first $t_0 - 1$ samples of $h(t)$ are assumed to correspond to the direct signal, $d(t)$, an energy ratio between the direct and reverberant components can be defined as [257, 262]:[5]

$$\frac{E_d}{E_r} = 10 \log \left\{ \frac{\sum_{t=0}^{t_0-1} d^2(t)}{\sum_{t_0}^{\infty} g^2(t)} \right\} \quad \text{(dB)} \tag{6.5a}$$

[3] *Early* reflections can be beneficial in that it subjectively reinforces the direct sound component. This reinforcement is what makes it easier to hold conversation in closed rooms compared with outdoors – especially, for example, in deep snow where there are no *early reflections* at all [201].

[4] It is implicitly assumed that the direct path component arrives at the source at $t = 0$.

[5] An equivalent criterion can also be established for continuous room impulse response by replacing the sum by an integral.

Therefore, the larger the ratio E_d/E_r, the less reverberation there is in the acoustic environment and, therefore, the higher the intelligibility of speech recorded in such an environment. If the impulse response is truncated, the infinite summation in the denominator of this expression is simply replaced by a finite summation [257].

The second criterion employed for the tests is designed to incorporate the *early portion* of the room impulse response which is actually beneficial to intelligibility [52, 201]. This criterion, often referred to as the 'early' to 'late' energy ratio [111, 201, 257], is defined as:[6]

$$\frac{E_e}{E_l} = 10 \log \left\{ \frac{\sum_{t=0}^{t_{50}-1} d^2(t)}{\sum_{t_{50}}^{\infty} g^2(t)} \right\} \quad \text{(dB)} \tag{6.5b}$$

where t_{50} is the sample corresponding to 50 milliseconds of continuous time data. This ratio is also known as the *clarity index*. The time-span of the early energy is sometimes set to 80 msec [111]. Finally, it is noted that Tokuyama [361] presents results which suggests the cross-correlation between undistorted and reverberant speech is effective as an objective method for estimating speech interference when compared with subjective measurements.

6.1.3.3 Reverberation Time

Reverberation time is the time interval, T_{60}, in which the reverberating sound energy, due to decaying reflections, reaches one millionth of its initial value, *i.e.* the time interval it takes for the reverberation level to drop by 60 dB [191, 201]. A number of reverberation formula exist;[7] for example, the relationship between the reverberation time, T_{60}, and the average damping constant, $\bar{\delta}$, of the resonant modes in the room is given by:[8]

$$T_{60} = \frac{6.91}{\bar{\delta}} \tag{6.6}$$

[6] The criteria in equations (6.5) are, in fact, simply signal-to-noise ratios, where the signal is the 'direct' component, and the noise is the 'reverberant' component.

[7] It is important to realise that reverberation time is actually a function of the excitation frequencies. For example, the reverberation time formula of (6.6) is derived using wave acoustics which, as noted in section §6.1.1.2, is only valid for frequencies above the first resonant mode.

[8] It is assumed that the damping constants of these resonant modes are approximately uniformly distributed, and $\bar{\delta}$ is a weighted average of these constants.

This expression is used in section §6.1.1.2 to derive a useful expression for the *Schroeder cut-off frequency* in equation (6.1). Typical reverberation times, usually measured at an excitation of 500 Hz, vary from around 0.3 seconds for living rooms, up to 10 seconds for large churches and reverberation chambers. Acoustic environments in which dereverberation is of interest for improving speech intelligibility include large rooms, where reverberation times lie between 0.7 and 2 seconds. Plomp and Duquesnoy [290] investigate the required reduction in reverberation time of various acoustic environments in order for speech in noise to be of acceptable intelligibility for the hearing-impaired.

6.1.3.4 Reverberation Distance

If a *diffuse* sound source continuously supplies acoustic energy into a room, there will be stationary sound energy throughout the room. Since reverberation can be considered as a statistical process with reflections off a large number of surfaces across the entire room, the energy density throughout the room due to reverberation will be constant. This is since, with a diffuse sound source, there is no *statistical* reason why the reverberant energy in one part of a closed room should be higher than that in another part, save exceptional cases, due to the large number of reflections; the room is said to be filled with a *diffuse sound field* [191]. However, since a spherical sound wave is attenuated with distance as a result of the spherical spread of sound pressure, the energy density due to direct sound between the source and a point in the room falls off with distance from the source, as shown in Figure 6.1 [201]. The distance at which the steady state reverberant energy equals the direct sound energy is called the *reverberation distance* or *radius* and, for a point sound source, is given by [201]:[9]

$$r_d = 0.1\sqrt{\frac{V}{\pi T_{60}}} \tag{6.7}$$

where V and T_{60} are, respectively, the volume (in m^3) and the reverberation time (in seconds) of the room. If an observer is within the reverberation distance of a source, the direct energy is greater than the reverberant energy while, if the observer is outside the reverberation distance, the reverberant energy will be dominant. The intelligibility of speech, for example, then depends greatly on whether the observer

[9] For other sound sources, the reverberation distance should be multiplied by the directivity factor of the source, *i.e.* the intensity generated in the direction under consideration divided by the average intensity.

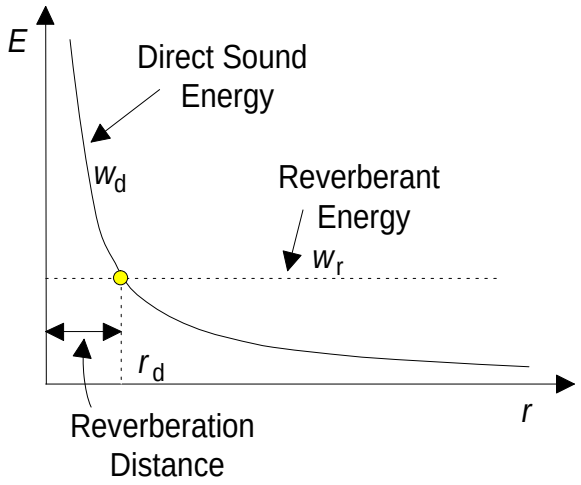


Figure 6.1: *Spatial dependence of direct and reverberant energy densities, w_d and w_r, respectively.*

is near the source, or far from the source, and explains why reverberation has negligible effect on intelligibility when using normal telephones or equipment where the microphone can be placed near to the sound source. In a typical small office of volume 40 m^3, for example, with $T_{60} = 0.5$ sec, the reverberation distance is 0.5 m.

6.1.4 Nonminimum-Phase Property

Room transfer functions are often nonminimum-phase because there is more energy in the reverberant component of the room impulse response than in the component corresponding to sound travelling along a direct path.[10] This property can be demonstrated by the simple example of sound waves travelling in a cylindrical tube of infinite length. Suppose the sound source emits a spherical sound wave, and consider only waves which arrive at the observation point. Figure 6.2 shows the path of the waves which arrive at the observation point after only a few re-

[10] It can be shown [281] that for a system with impulse response $h(t)$,

$$\sum_{t=0}^{\tau} |h(t)|^2 \leq \sum_{t=0}^{\tau} |h_{min}(t)|^2 \ , \quad \forall \tau \geq 0$$

where $h_{min}(t)$ is the minimum-phase system corresponding to $h(t)$. From Parseval's Theorem, equality results as $\tau \to \infty$ since $h(t)$ and $h_{min}(t)$ have identical magnitude frequency response. This result implies that the impulse response of the minimum-phase system, $h_{min}(t)$, is more compressed towards the origin than $h(t)$ or, alternatively, that $h_{min}(t)$ is delayed the least [281, 353]. Thus, since the reverberant component of an AIR generally contains more energy than the direct component, the impulse response is, in general, nonminimum-phase.

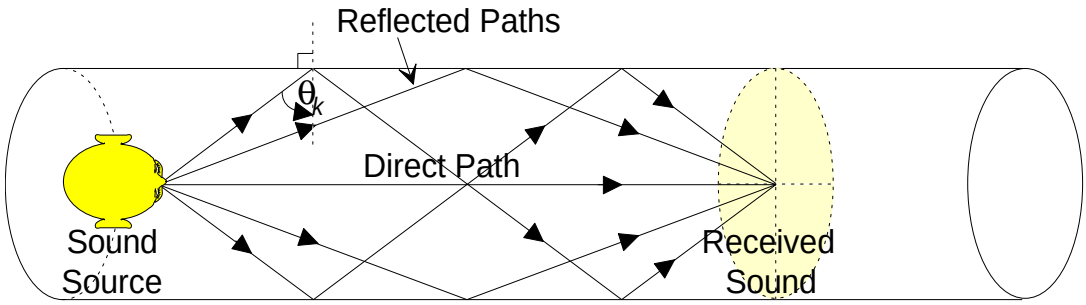


Figure 6.2: *In an infinitely long cylindrical tube, the reverberant energy is greater than the energy contained in the sound travelling along a direct path, thus demonstrating the nonminimum-phase properties of room acoustics.*

flections. If the tube-walls have a complex reflection coefficient, $R(\theta)$, where θ is the angle between the wall normal and the direction of wave propagation, then the wall absorbs the fraction[11] $\alpha = 1 - |R(\theta)|^2$ of the incident energy during the waves reflection. It may be shown that the ratio of the energy of all the reflected waves at the observation point to the energy of the direct wave is given by:

$$\eta = 2\pi r^2 \sum_{k=1}^{\infty} \frac{|R(\theta_k)|^{2k}}{d_k} \quad \text{where } R(\theta_k) = \frac{\zeta \cos\theta_k - 1}{\zeta \cos\theta_k + 1},\ \cos\theta_k = \frac{2kr}{d_k},$$

ζ is the *specific acoustic impedance* of the wall, θ_i is the angle of incidence of the plain waves, r and L are the radius and length of the tube respectively, and $d_k = \sqrt{L^2 + (2kr)^2}$ is the distance the wave travels from source to observation point. The attenuation factor, d_k, in the summation arises due to the decay in amplitude with distance of spherical sound waves as a result of the spherical spread of the sound pressure generated by the source. There is clearly a range of values of η for which the reverberant energy is greater than the direct energy and, therefore, when the RTF is nonminimum-phase.

6.1.5 Room Impulse Response Measurement

Kuttruff [201] discusses, in detail, practical measurement techniques for room acoustics. The acoustic impulse response (AIR) is the main acoustical property of interest in dereverberation, and its measurement can be considered as a system identification problem. Acoustic impulse responses can be acquired using white Gaussian

[11] This fraction is called the *absorption coefficient of the wall* [201].

noise (WGN) sequences [201,353], binary pseudo-random sequences [106,201], or frequency sweeps [34]. Impulsive excitations are usually avoided since they are always approximated by short finite pulses and, furthermore, in order to attain a given signal-to-noise ratio (SNR), the energy of the excitation must exceed a limit which could cause the device creating the excitation, such as a loudspeaker, damage or to cease operating in its linear region. Frequency sweeps are used when the impulse response must be measured in a short time, whereas long WGN sequences are required in order to attain a reasonable SNR. Finally, it is essential that the distance between the sound source and the measuring microphone is greater than the *reverberation distance*, otherwise the *reverberant sound field* in question is partially masked by the direct sound field of the source.

The AIRs, $h(t)$, used throughout this work are measured from the response to a WGN sequence using the unbiased cross-correlation method for system identification: *i.e.* $h(t) = R_{ye}(t)\,,\ t \in \mathscr{L} = \{-L, \ldots, L\}$, where $\{e(t), t \in \mathscr{T}\}$, $\mathscr{T} = \{0, \ldots, T\}$, is a random WGN excitation sequence, $\{y(t),\ t \in \mathscr{T}\}$ is the observed response, and $\{h(t),\ t \in \mathscr{L}\}$ is the AIR.[12] The cross-correlation is calculated using the *unbiased* form of the sample correlation function [353]:

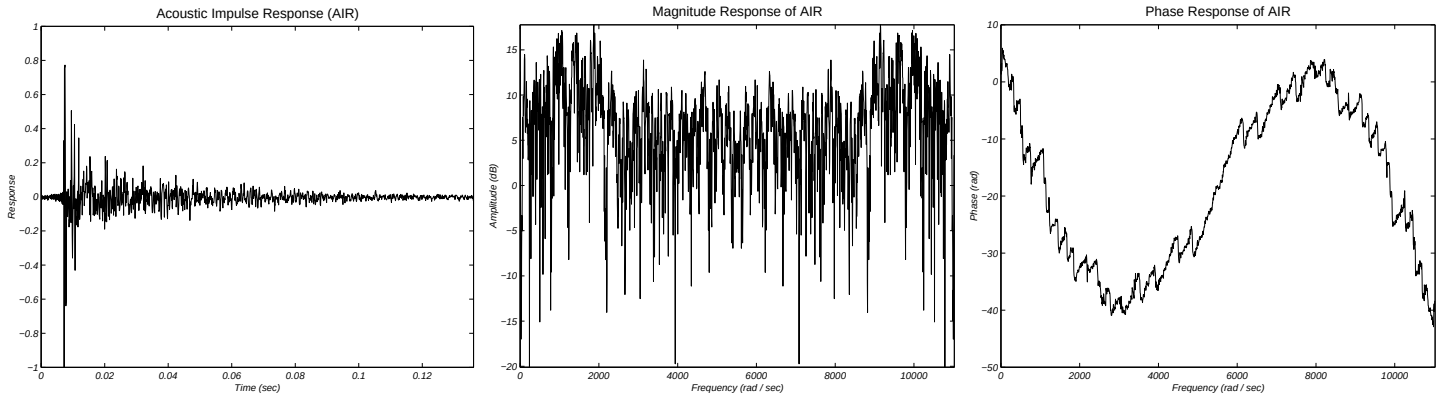
$$\hat{R}_{ye}(t) = \begin{cases} \dfrac{1}{T-t} \displaystyle\sum_{\tau=0}^{T-1-t} y(\tau+t)\, e(\tau), & \text{if } 0 \le t < T, \\ \dfrac{1}{T-|t|} \displaystyle\sum_{\tau=0}^{T-1-|t|} y(\tau)\, e(\tau+|t|), & \text{if } -T < t \le 0 \end{cases} \tag{6.8a}$$

Since $y(\tau + t) = 0$ if $\tau > T - 1 - t$ in the first expression, and $e^*(\tau + |t|) = 0$ if $\tau > T - 1 - |t|$ in the second, the limits in (6.8a) may be set to $T - 1$. Hence:

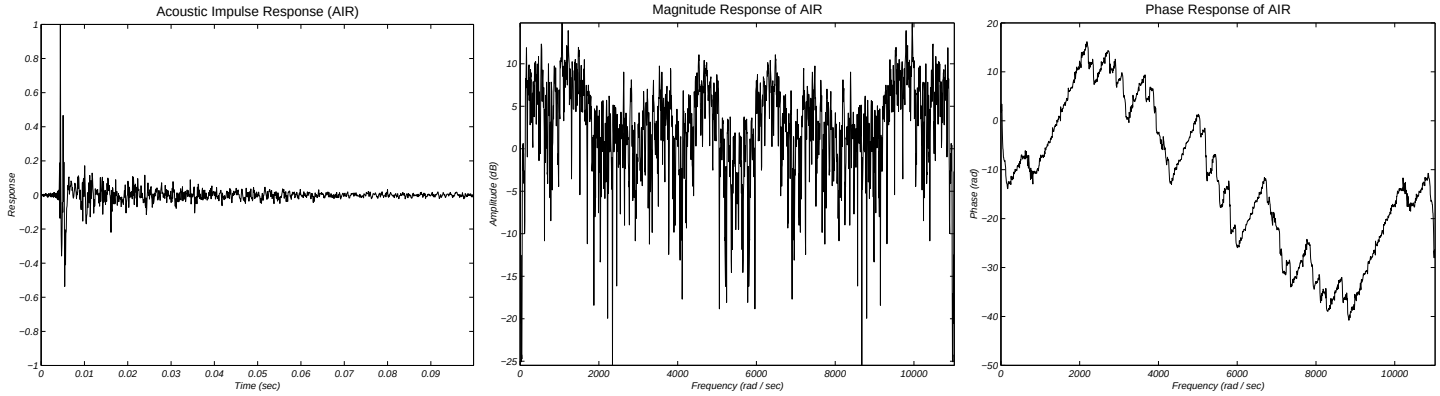
$$\hat{R}_{ye}(t) = \begin{cases} \dfrac{1}{T-t}\, \mathscr{F}^{-1}\{Y(k)\, E^*(k)\} & \text{if } 0 \le t < T, \\ \dfrac{1}{T-|t|}\, \mathscr{F}^{-1}\{Y^*(k)\, E(k)\} & \text{if } -T < t \le 0 \end{cases} \tag{6.8b}$$

where $y(\tau) \rightleftharpoons Y(k)$ and $e(\tau) \rightleftharpoons E(k)$ are discrete Fourier transforms (DFT), and $\mathscr{F}^{-1}\{Z(k)\}$ denotes the inverse DFT of $Z(k)$ [382]. A second form of the sample correlation function is identical to equations (6.8) except the scaling term is simply $1/T$; this is known as the *biased* form of the sample correlation function, and is useful when the amount of data available for the correlation estimate is small [353].

[12] Note that $L \ll T$, since the excitation sequence is assumed to be much longer than the length of the acoustic impulse response of the room.

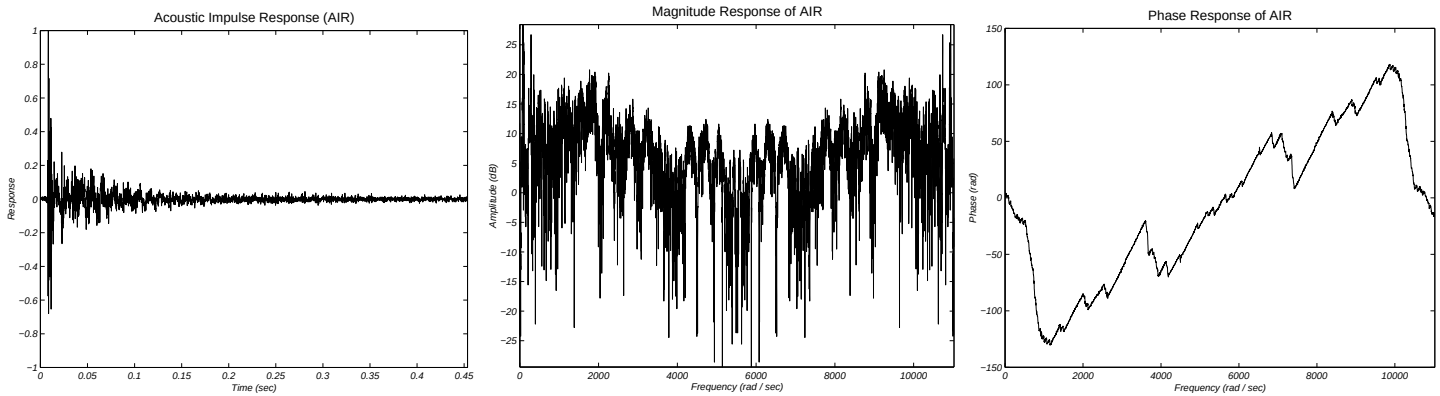


(c) Source and observer in an office, distances typical of interview or teleconferencing scenarios.

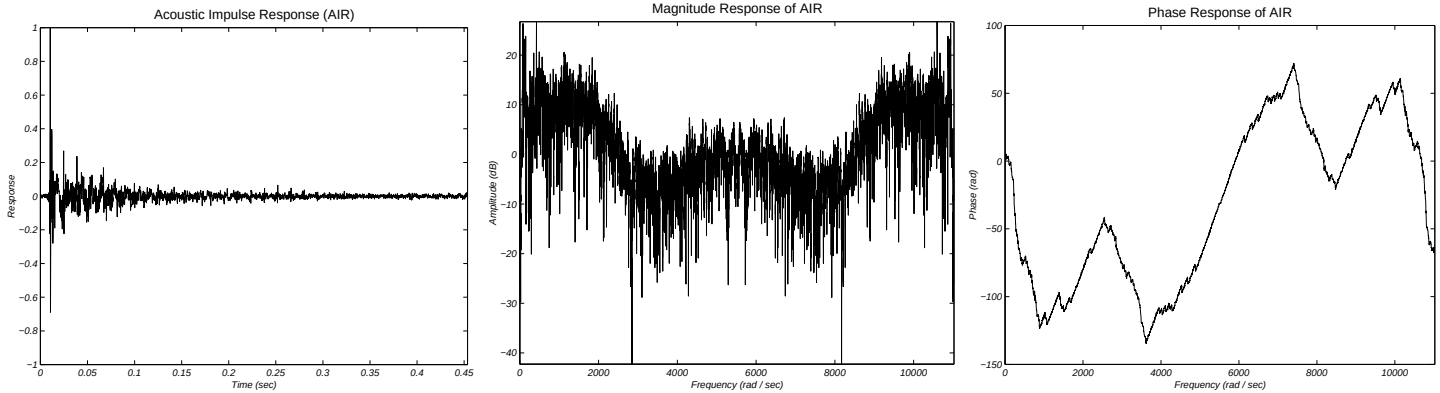


(d) Source and observer above a large hard surface, and distant from other boundaries.

Figure 6.3: *Typical impulse responses of a number of different acoustic environments.*



(a) Source and observer in a stairwell with a direct line-of-sight path.



(b) Source and observer in a stairwell with no direct path.

6.1.6 Typical acoustic impulse responses

Figure 6.3 shows the impulse and frequency responses of some typical acoustic impulse responses, measured using the approach discussed in section §6.1.5. These responses are representative of a number of different acoustic environments, and include the source and observer:

Figure 6.3(a): in a stairwell, with a direct line-of-sight path from source to observer.

Figure 6.3(b): in a stairwell, with no direct path from source to observer.

Figure 6.3(c): in a typical office, distances typical of teleconferencing scenarios.

Figure 6.3(d): above a large hard surface, and distant from other boundaries.

6.2 Modelling of Room Transfer Functions

The ultimate aim in this dissertation is to *dereverberate* distorted speech recorded in an echoic acoustic environment and, to achieve this, the acoustical properties of the room must be modelled. This section discusses the suitability of some well-known modelling techniques in signal processing for the representation of room acoustics, the robustness of the models to variations in the source and observer position, and the effect of parameter variation on the accuracy of the model. These models are reintroduced from a signal processing perspective in section §7.4. Although this dissertation is primarily concerned with *modelling* the impulse response of a real room, it is instructive to consider techniques which *simulate* the impulse response of a room, and the most well known of these simulation techniques is discussed in the next section.

6.2.1 Image Method for Simulating Room Acoustics

By far the most well known technique for *simulating* the impulse response of a room is the *image method* [8, 201, 354], which essentially sums all the initial, or first-order, reflections of the sound field within a room with the resulting higher order reflections using the assumption of *geometrical room acoustics*. A point sound

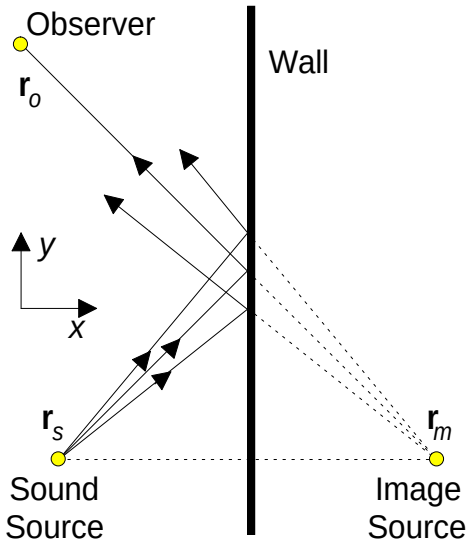


Figure 6.4: *Construction of a mirror source.*

source of single frequency ω, at position $\mathbf{r}_s = (x, y, z)$ in free space, emits a pressure wave $P_{(\mathbf{r}_s,\mathbf{r}_o)}(\omega, t)$ at position $\mathbf{r}_0 = (x', y', z')$ of the form [191, 201]:

$$P_{(\mathbf{r}_s,\mathbf{r}_o)}(\omega, t) = P_0 \frac{\exp\left[j\omega(r/c - t)\right]}{r} \tag{6.9}$$

where c is the speed of sound, $t \in \mathcal{T} \subset \mathbb{R}$ is time, and $r = |\mathbf{r}_s - \mathbf{r}_0|$. The reflection of *sound rays* from the source at $\mathbf{r}_s$ can be illustrated by placing an *image source* symmetrically on the far side of the wall,[13] provided that the reflecting surface is plane (see Figure 6.4). Consider the sound transmission from the source $\mathbf{r}_s$ to the observation point $\mathbf{r}_o$ in free space with one reflecting wall near by. The sound travels along the direct path and by reflection from the wall. The path of the reflected ray is identical to sound travelling along a direct path from the image source when the wall is disregarded, provided the power spectrum of the image source is modified to account for the frequency dependent absorption of the wall. For example, if the wall is rigid,[14] the total pressure wave at the observation point can be written as

$$\hat{P}_{(\mathbf{r}_s,\mathbf{r}_o)}(\omega, t) = P_{(\mathbf{r}_s,\mathbf{r}_o)}(\omega, t) + P_{(\mathbf{r}_m,\mathbf{r}_o)}(\omega, t) \tag{6.10}$$

[13] In accordance with Snell's law, when an acoustic wave propagating in an isotropic medium strikes a wall, the resulting reflection angle is equal to the incident angle, if the 'roughness' dimensions of the wall are small compared to the acoustic wavelength [168].

[14] A rigid wall has perfect reflection, so its reflection factor $|R(\theta)| = 1$.

where $P_{(\cdot,\cdot)}(\omega, t)$ is given by equation (6.9). In the more general case of six rigid walls, the situation becomes more complex because each image is itself an image. The pressure at the observation point can then be written as [8] (see [195, 354] for further results):[15]

$$\hat{P}_{(\mathbf{r}_s,\mathbf{r}_0)}(\omega, t) = P_0 \sum_{\mathbf{p}=0}^{1} \sum_{\mathbf{q}=-\infty}^{\infty} \frac{\exp\left[j\omega\left|\mathbf{r_p} + \mathbf{r_q}\right|/c - t\right]}{\left|\mathbf{r_p} + \mathbf{r}_q\right|} \tag{6.11}$$

where $\mathbf{r}_p$ is expressed in terms of the integer 3-vector $\mathbf{p} = (u, v, w)$ as

$$\mathbf{r}_p = (x + (2u-1)x', y + (2v-1)y', z + (2w-1)z') \tag{6.12}$$

which is equivalent to the eight vectors given by the eight permutations over $\pm$ of $(x \pm x', y \pm y', z \pm z')$. Further, $\mathbf{r_q} = 2(lL_x, mL_y, nL_z)$, where $\mathbf{q} = (l, m, n)$ is an integer 3-vector, and (L_x, L_y, L_z) are the room dimensions. The AIR is, therefore,

$$h_{(\mathbf{r}_s,\mathbf{r}_0)}(t) = P_0 \sum_{\mathbf{p}=0}^{1} \sum_{\mathbf{q}=-\infty}^{\infty} \frac{\delta\left(t - \left|\mathbf{r_p} + \mathbf{r_q}\right|/c\right)}{\left|\mathbf{r_p} + \mathbf{r_q}\right|} \tag{6.13a}$$

This expression can also be derived directly from the wave equation as shown in [8]. Furthermore, using the same approximate model for non-rigid walls, assuming *an angle independent specific acoustic impedance*,[16] denoted by ζ_k, it may be shown that the modified room impulse response is given by:

$$h'_{(\mathbf{r}_s,\mathbf{r}_0)}(t) = P_0 \sum_{\mathbf{p}=0}^{1} \sum_{\mathbf{q}=-\infty}^{\infty} \zeta_{x1}^{|l-u|}\zeta_{x2}^{|l|}\,\zeta_{y1}^{|m-v|}\zeta_{y2}^{|m|}\,\zeta_{z1}^{|n-w|}\zeta_{z2}^{|n|} \frac{\delta\left(t - \left|\mathbf{r_p} + \mathbf{r_q}\right|/c\right)}{\left|\mathbf{r_p} + \mathbf{r_q}\right|} \tag{6.13b}$$

Note that both equations (6.13a) and (6.13b) may be expressed in the general form

$$h'_{(\mathbf{r}_s,\mathbf{r}_0)}(t) = P_0 \sum_{p=-\infty}^{\infty} h_q\,\delta\left(t - \tau_q\right) \tag{6.13c}$$

again providing justification for modelling room acoustics using LTI representations. This method can be combined with the *ray tracing method* to simulate room transfer functions in the medium frequency range where neither geometric acoustics, nor wave acoustics, are suitable [112, 168]. Champagne *et al.* [59] use the image method to simulate a moving speaker in a rectangular room.

[15] Note that this representation assumes a corner of the room coincides with the origin, (0,0,0).

[16] For a complete discussion of the assumptions used in this model, see [8].

6.2.2 Pole-Zero Modelling

Since the RTF (6.3) can be expressed by a rational expression, it can be modelled by the conventional pole-zero model[17] with poles $\{p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\, i \in \mathscr{P}\}$, $\mathscr{P} = \{1,\ldots,P\}$ and zeros $\{q^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\, i \in \mathscr{Q}\}$, $\mathscr{Q} = \{1,\ldots,Q\}$, or AR coefficients $\{a^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\, i \in \mathscr{P}\}$ and moving average (MA) coefficients $\{b^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\, i \in \hat{\mathscr{Q}}\}$, $\hat{\mathscr{Q}} = \{1,\ldots,Q+R\}$:

$$H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = C^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}\, z^{R}\, \frac{\prod_{i=1}^{Q}\left[1 - q^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]}{\prod_{i=1}^{P}\left[1 - p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]} \equiv \frac{\sum_{i=1}^{Q+R} b^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-i}}{1 + \sum_{i=1}^{P} a^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-i}} \tag{6.14a}$$

where $H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z)$ represents the pole-zero modelled RTF, P is the number of poles, $Q + R$ is the total number of zeros including those at the origin, and $C^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}$ is a gain constant. Since most room transfer functions are *stable* and *causal*, the denominator of this transfer function must correspond to a stable causal sequence and, therefore, the poles must lie within the unit circle: $\left|p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\right| < 1,\, i \in \mathscr{P}$.[18] However, as discussed in section §6.1.4, AIRs are often nonminimum-phase, and so the zeros $q^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\, i \in \mathscr{Q}$ may lie outside the unit circle. Alternatively, equation (6.14a) may be expressed as:

$$H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = C^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}\, z^{R}\, \frac{\prod_{i=1}^{Q_m}\left[1 - r^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right] \prod_{i=1}^{Q_n}\left[1 - s^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z\right]}{\prod_{i=1}^{P}\left[1 - p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]} \tag{6.14b}$$

where $\left|r^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\right| < 1,\, i \in \mathscr{Q}_m$, and the zeros $\left|s^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\right| < 1,\, i \in \mathscr{Q}_n$, correspond to the nonminimum-phase component of the transfer function. Although the distribution of nonminimum-phase zeros in the *complex frequency plane*,[19] given the location of the RTF poles, is discussed in the literature [231, 232, 356–358], these results do not aid the prediction of the general distribution of zeros in the $\mathscr{Z}$-

[17] See section §7.4 for further details on pole-zero modelling.

[18] A nonminimum exponentially stable AR process is noncausal and possesses the transfer function

$$H(z) = \frac{1}{\prod_{i=1}^{P_1}\left[1 - p_1(i)\, z^{-1}\right] \prod_{i=1}^{P_2}\left[1 - p_2(i)\, z\right]}$$

Although this gives rise to a more general form of equation (6.14a), the resulting system would be noncausal and, as such, be too general to model a causal room response.

[19] Points on the complex frequency plane are obtained from the frequency response of a system (*i.e.* equation (6.3)) by substituting a complex frequency, $\omega_r + j\omega_i$, for the real frequency, ω.

domain for use in the inversion of the AIR. Bistritz [40, 41] discusses the problem of determining the location of zeros w. r. t. the unit circle, which may be useful for determining the number of nonminimum-phase zeros. Mourjopoulos and Paraskevas [262] discuss, in detail, pole-zero modelling of RTFs, and the model has often been used in the literature, for example [262, 268]. From a physical point of view, poles represent resonances (see equation (6.3)), and zeros represent time delays and anti-resonances. The characteristics of all-zero and all-pole models when used to represent room acoustics are described in the sections §6.2.4 and §6.2.5.

6.2.3 Pole-Zero Model Decompositions

There are two useful decompositions of equation (6.14b) which are useful for inverting room transfer functions. The first is to write (6.14b) as [256, 259]:

$$H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = H^{PZ\ min}_{(\mathbf{r}_s,\mathbf{r}_0)}(z)\ H^{PZ\ max}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \tag{6.15a}$$

where the minimum and maximum-phase components are given by:

$$H^{PZ\ min}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \triangleq C^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}\, z^R\, \frac{\prod_{i=1}^{Q_m}\left[1 - r^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]}{\prod_{i=1}^{P}\left[1 - p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]} \tag{6.15b}$$

and

$$H^{PZ\ max}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \triangleq \prod_{i=1}^{Q_n}\left[1 - s^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z\right] \tag{6.15c}$$

The second is to observe that equation (6.14b) can also be decomposed into its *equivalent minimum-phase* function and a nonminimum-phase *all-pass* function [281, Chapter 7], [140, 256, 257, 259, 262, 274, 353]:

$$H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = H^{PZ\,mp}_{(\mathbf{r}_s,\mathbf{r}_0)}(z)\ H^{PZ\,ap}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \tag{6.16a}$$

where the *equivalent minimum-phase* component of the RTF is given by:

$$H^{PZ\,mp}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = C^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}\, z^R\, \frac{\prod_{i=1}^{Q_m}\left[1 - r^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]\prod_{i=1}^{Q_n}\left[1 - s^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]}{\prod_{i=1}^{P}\left[1 - p^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]} \tag{6.16b}$$

and $H^{PZ\,ap}_{(r_s,r_0)}(z)$ is the *all-pass* component by:

$$H^{PZ\,ap}_{(r_s,r_0)}(z) = \frac{\prod_{i=1}^{Q_n}\left[1 - s^{PZ}_{(r_s,r_0)}(i)\,z\right]}{\prod_{i=1}^{Q_n}\left[1 - s^{PZ}_{(r_s,r_0)}(i)\,z^{-1}\right]} \tag{6.16c}$$

where $\left|H^{PZ\,ap}_{(r_s,r_0)}(z)\right| = 1$, for $z = e^{j\omega}$, $\forall\omega$.

6.2.4 All-zero RTF Model

The RTF of equation (6.3) can be modelled by the conventional all-zero model, or finite impulse response (FIR) filter, which can be represented with either zeros $\{q^Z_{(r_s,r_0)}(i),\, i \in \mathcal{Q}\}$, $\mathcal{Q} = \{1,\ldots,Q\}$, or MA coefficients $\{b^Z_{(r_s,r_0)}(i),\, i \in \hat{\mathcal{Q}}\}$, $\hat{\mathcal{Q}} = \{1,\ldots,Q+R\}$, and can effectively be considered as the numerator of equation (6.14a):

$$H^Z_{(r_s,r_0)}(z) = C^Z_{(r_s,r_0)}\,z^R \prod_{i=1}^{Q}\left[1 - q^Z_{(r_s,r_0)}(i)\,z^{-1}\right] = \sum_{i=1}^{Q+R} b^Z_{(r_s,r_0)}(i)\,z^{-i} \tag{6.17a}$$

As discussed in the previous section, this can also be expressed in the form:

$$H^Z_{(r_s,r_0)}(z) = C^Z_{(r_s,r_0)}\,z^R \prod_{i=1}^{Q_m}\left[1 - r^Z_{(r_s,r_0)}(i)\,z^{-1}\right]\prod_{i=1}^{Q_n}\left[1 - s^Z_{(r_s,r_0)}(i)\,z\right] \tag{6.17b}$$

where $\left|r^Z_{(r_s,r_0)}(i)\right| < 1$, $i \in \mathcal{Q}_m$ and $\left|s^Z_{(r_s,r_0)}(i)\right| < 1$, $i \in \mathcal{Q}_n$. The first product term in (6.17b) corresponds to the minimum-phase component, and the second product term corresponds to the maximum-phase component of the all-zero RTF. This expansion is used in section §6.3.3 when inversion of the RTF is considered. There are several main limitations of FIR filters imposed by the nature of room acoustics [257–259, 262]:

1. Room impulse responses are, in general, very long and an all-zero filter typically requires up to 10,000 coefficients, approximately determined by

$$n_s = T_{60} \cdot f_s \quad \text{(samples)} \tag{6.18}$$

where f_s is the sampling frequency in Hertz, and T_{60} is the reverberation

time for the enclosure, dictating a 60-dB dynamic range, as discussed in section §6.1.3.3. As an example, if $T_{60} = 0.5$ seconds and $f_s = 44.1$ kHz, $n_s = 22050$ samples [259]. In, for instance, acoustic echo cancellation, long reverberation times accentuate inverse filtering problems due to the large number of taps required [158].

2. The resulting FIR filter may be effective and appropriate only for a very limited spatial combination of source and receiver positions within a particular enclosure [144]. As discussed in section §6.3.3, the large variations in RTF for small changes in source–observer positions can actually cause invertibility problems and, in some cases, the distortion of the 'equalised' transfer function will be greater than the original distortion due to the RTF [80, 257–259, 305]. This sensitivity can be explained by the nature of room acoustics; transfer function zeros result from local cancellations of multipath sound components which are easily disturbed by slight changes in source–observer positions [262]. This suggests that if the room impulse response is incorrectly estimated there may be problems with equalisation.

Kale *et al.* [173, 233] use a state-space technique known as balanced model truncation (BMT) [30] to convert a high-order FIR filter into a reduced-order IIR filter to model both the transfer characteristics of the AIR of a motor car [173], and the *head-related transfer function* (HRTF) [233] (see page 139 and footnote 21). BMT delivers a faithful reproduction of the phase characteristics of the original FIR response which may have been lost if conventional least-squares ARMA or AR modelling methods were used. Additionally, Tohyama and Lyon [356] discuss the effect of truncating an impulse response and demonstrate, for example, that truncation can change the minimum-phase behaviour of a RTF into a nonminimum-phase characteristic.

6.2.5 All-pole RTF Model

An alternative to equation (6.17a) for the representation of the of (6.3) is the causal all-pole model, or infinite impulse response (IIR) filter, which can be represented either by the poles $\{p^P_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\ i \in \mathscr{P}\}$, $\mathscr{P} = \{1, \ldots, P\}$, or AR coefficients $\{a^P_{(\mathbf{r}_s,\mathbf{r}_0)}(i),\ i \in \mathscr{P}\}$, and can effectively be considered as the denominator of equa-

tion (6.14a):

$$H^{P}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = \frac{C^{P}_{(\mathbf{r}_s,\mathbf{r}_0)}}{\prod_{i=1}^{P}\left[1 - p^{P}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right]} \equiv \frac{C^{P}_{(\mathbf{r}_s,\mathbf{r}_0)}}{1 + \sum_{i=1}^{P} a^{P}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-i}} \tag{6.19}$$

The all-pole or autoregressive model for approximating rational transfer functions is widely used in many fields, especially in speech analysis [237], and this application is discussed in more detail in section §7.6. Typical all-pole model orders required for approximating room transfer functions are in the range $50 \leq P \leq 500$ [262], compared to speech analysis applications when the model is usually much lower, typically $4 \leq p \leq 30$ [237]; a theoretical all-pole model order is discussed in section §6.2.7. Mourjopoulos and Paraskevas [262] state that all-pole model orders are typically a factor of 40 lower than all-zero model orders, while several studies by Gudvangen and Flockton [132, 133] state that the gain achieved using pole-zero over all-zero modelling of reverberant acoustic environments is not as high as generally thought throughout the literature, with reduction in coefficients typically in the order of 1.2 to 1.5. These latter studies use modelling error functions to measure the fit of the pole-zero models to the complete AIR, rather than fitting the most important reverberant characteristics. Therefore, a significant reduction in model order should be expected for many applications where it is more important to model the main reverberant component, as discussed in section §6.1.3, rather than just minimising the modelling error.

A significant advantage of the all-pole model over the all-zero model is its lower sensitivity to changes in source and observer positions, a property which is taken advantage of, and discussed further, in section §6.2.6. Mourjopoulos and Paraskevas [262] conclude that in many signal processing applications dealing with room acoustics, it may be both sufficient and more efficient to manipulate all-pole model coefficients rather than high-order all-zero models. The all-pole model is also the basis of the technique discussed in [263] which allows the classification of all possible RTFs corresponding to different source–observation positions, thereby providing a 'codebook' for possible transmission paths in dereverberation applications. A shortcoming of the causal all-pole filter is that, since it is causal and stable, it is minimum-phase and, therefore, cannot model the nonminimum-phase component of room acoustics (section §6.1.4). Nevertheless, a subband all-pole model is introduced in Chapter 9 which avoids this problem since only a number of subbands considered individually have nonminimum-phase characteristics [384, 386].

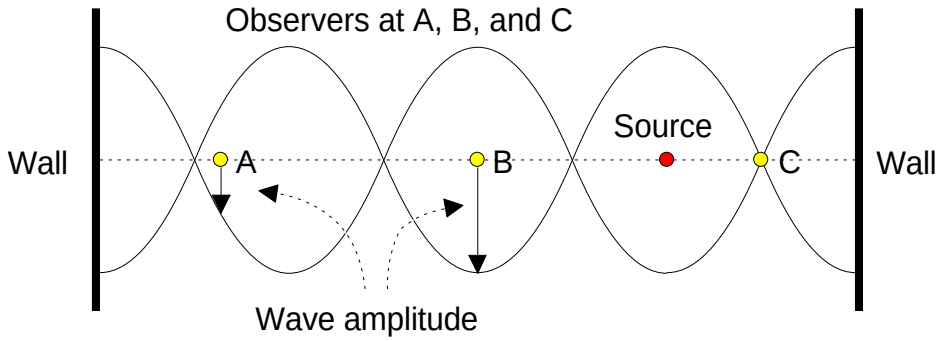


Figure 6.5: *Standing waves occur at resonances as shown for this 1-D room, and can be observed at any point in the room except node points (such as point C). Since this standing wave occurs independently of the source location and can be observed at all observation points, the acoustical poles which reflect the information of the resonant frequencies are independent of source–observer locations.*

Subband methods also reduce model complexity and the errors generated when full-bandwidth modelling is attempted [375].

6.2.6 Common Acoustical Pole and Zero Modelling

The room transfer function derivation using wave acoustics, as discussed in sections §6.1.1.2 and §6.1.2, leads to the form given in equation (6.3) which can be rewritten in the Laplace transform domain (s-plane) as:

$$H_{(\mathbf{r}_s,\mathbf{r}_o)}(s) = \prod_{i=1}^{\infty} \frac{B_{(\mathbf{r}_s,\mathbf{r}_o)}(s)}{(s - s_i)(s + s_i)} \tag{6.20}$$

for some polynomial $B_{(\mathbf{r}_s,\mathbf{r}_o)}(s)$ in s, where $s = j\omega$ and $s_i = j\omega_i - \delta_i$; if $\delta_i \ll \omega_i$, the resonances of the RTF occur at $\omega = \omega_i$. Acoustical poles are approximately independent of the source and observer position, which equations (6.14a) and (6.19) assumed was not the case, since they correspond to the resonant frequencies of a room; *standing waves* occur at these resonances, and can be observed at any point in the room, except at node points, as depicted in the 1-D case shown in Figure 6.5. Therefore, since this standing wave occurs independently of the source location, and can be observed at all observation points, the acoustical poles reflecting the information of the resonant frequencies must be independent of source–observer locations. The amplitude of the standing wave varies depending on the receiver positions, as shown in Figure 6.5, and this variation is reflected in the zeros of the

RTF [144]. As such, equation (6.14b) can be written in the simpler form:[20]

$$H^{\text{CAPZ}}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = C^{\text{CAPZ}}_{(\mathbf{r}_s,\mathbf{r}_0)}\, z^{R}\, \frac{\prod_{i=1}^{Q_m}\left[1 - r^{\text{CAPZ}}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z^{-1}\right] \prod_{i=1}^{Q_n}\left[1 - s^{\text{CAPZ}}_{(\mathbf{r}_s,\mathbf{r}_0)}(i)\, z\right]}{\prod_{i=1}^{P}\left[1 - p^{\text{CAPZ}}(i)\, z^{-1}\right]} \tag{6.21}$$

where $\{p^{\text{CAPZ}}(i),\, i \in \mathscr{P}\}$ are the common-acoustical poles independent of $(\mathbf{r}_s, \mathbf{r}_0)$. Nevertheless, it should be noted that the acoustical argument used for the justification of the common-acoustical pole and zero (CAPZ) model is simplistic, and other investigations on the fluctuations of AIRs within reverberant environments suggest that this assumption may not be strictly true [305]. The expression in equation (6.21) is known as the CAPZ model of RTFs, and was first introduced by Haneda *et al.* [143, 144] and recently extended [142]. The CAPZ model is particularly useful in applications where multiple room transfer functions from different source–observer positions are modelled, which could have applications in multi-channel source separation (for example, [119]). Recently the CAPZ model has been applied to the head-related transfer function (HRTF) [146] which describes the sound transmission characteristics from a source to a point in the ear canal in a *free field*.[21] The CAPZ model reiterates the observation that the all-pole model is more robust to changes in source–observer positions, and thus suggests using all-pole RTF models in singlel-channel dereverberation.

6.2.7 Theoretical Pole Order

The harmonic solutions of the wave-equation (6.2) for a rectangular room can be found using the separation of variables technique for solving partial differential equations. The number of modes in the harmonic solution can be counted, and it may be shown that the order of modes $N(f_u)$ for a room of dimensions L_x, L_y, L_z, with volume $V = L_x L_y L_z$, up to an upper frequency limit f_u is given by [201]:

$$N(f_u) = \frac{4\pi}{3} V \left(\frac{f_u}{c}\right)^3 + \frac{\pi}{4} S \left(\frac{f_u}{c}\right)^2 + \frac{L}{8}\left(\frac{f_u}{c}\right) \tag{6.22a}$$

where $S = 2\,(L_x L_y + L_x L_z + L_y L_z)$ is the room's surface area and $L = 4\,(L_x + L_y + L_z)$ is the sum of all the edge lengths occurring in the rectangular room.

[20] An equivalent form of equation (6.14a) can also be constructed.

[21] A *free field* means that there are no sound reflections from objects other than the human-head. The head-related transfer function only models the head, and not the acoustics of the room.

If $f_u \gg 500$ Hz, and $V \gg S$, then the last two terms in (6.22a) can be ignored. Furthermore, in the limiting case of $f_u \to \infty$, the first term corresponds to the number of modes for arbitrary shaped rooms, not just rectangular ones [201]. The order of the all-pole model up to a given sampling frequency, f_s, is therefore given by $P \approx 2N\,(f_s/2)$, or [144]:

$$P \approx \frac{\pi}{3} V \left(\frac{f_s}{c}\right)^3 \tag{6.22b}$$

If the all-pole model order is the same as the theoretical order in equation (6.22b), the all-pole model corresponds well with the actual room response. If the model order is lower than the theoretical order, the estimated poles correspond to the major resonance frequencies which have high Q factors [144]. This is typically the case since, for example, a typical small office with volume 40 m^3 has $\sim 1.75 \times 10^5$ acoustic modes with natural frequencies below 3.5 kHz giving a very high all-pole model order. Moreover, it is clear for most typical rooms that the model order given by equation (6.22b) is much greater than the typically length of an FIR filter, as given in equation (6.18). Hence, equation (6.22b) is a very high upper bound.

6.3 Dereverberation of Speech Signals

Reverberation reduction processes may generally be divided into single or multiple microphone methods, and into those primarily affecting coloration or those affecting reverberant tails. Early room echoes mainly contribute to coloration, or spectral distortion, while late echoes, or long term reverberation, contribute noise-like perceptions or tails on speech signals [9]. The following section reviews some existing approaches, in addition to those discussed in section §6.3.3, to the enhancement of reverberant speech.

6.3.1 Existing Approaches to Reverberant Speech Enhancement

Multiple-microphone techniques for speech enhancement applications, by removing long-term echo, date back to the work of Allen *et al.* in 1977 [9], in which sound recordings are made using two microphones, with the processing performed in the frequency domain. The process uses a subband method and, within each

band, the delay existing between the 'coherent part' of the two microphone signals (*i.e.* the actual signal and early echoes but not the late reflections) is removed by shifting the phase of one signal to align it with the other. These phase corrected signals are then summed and the entire process is referred to as *cophase and add* in bands. The gain of each band is then adjusted according to the normalised cross-correlation function of the observed signals. This has the effect of attenuating bands with low levels of 'coherence' containing mainly reverberation, while passing relatively unaltered, or slightly enhanced, frequency bands with a strong level of 'coherence' implying the presence of a strong direct component and early echoes. The final stage is to resynthesis the signal by combining the signals in each subband. Since that early study, further investigations into this technique have been undertaken and subsequent modifications suggested [42].

Dereverberation techniques based on microphone arrays is still a highly active area of research [219, 245]. However, single microphone reverberant speech enhancement requires some prior knowledge of the impulse response of the acoustic environment, although various schemes attempt to reduce the amount needed. Langhans and Strube [206] investigate multiband envelope filtering techniques for speech enhancement, and Mourjopoulos and Hammond [261] use a (nonlinear) multiband envelope convolution method which reduces the prior information required regarding the room transmission characteristics, and lessens the sensitivity of the processing method to inaccuracy in the model and measurements of the room impulse response. According to the multiband envelope convolution method, the envelope of the reverberant speech in each frequency band can be approximated by the convolution of the clean anechoic speech signal with the envelope of the acoustic impulse response. As such, the problem of enhancement reduces to the deconvolution of the room response envelope, and the reconstruction of the speech signal. This method reduces the effect of spectral coloration on the observed reverberant speech, but not the long-term effect of reverberation which is mostly contained in the *nonminimum excess phase* component of the signal, as discussed in section §6.3.2. A recent related approach by Hirobayashi *et al.* [155] uses power envelope inverse filtering; specifically, the original waveform and acoustic impulse response are modelled as:

$$s(t) = s_{env}(t)\, n_s(t) \tag{6.23a}$$

$$h(t) = h_{env}(t)\, n_h(t) \tag{6.23b}$$

where

$$h_{env}(t) = a\, e^{-\frac{6.9t}{T_{60}}} \tag{6.23c}$$

where $n_s(t)$ and $n_h(t)$ are zero mean, unit variance, white Gaussian noise processes. The power envelope, $x^2_{env}(t)$, of the reverberant speech, $x(t)$, is then given by

$$E\left[y^2(t)\right] = x^2_{env}(t) = s^2_{env}(t) \star h^2_{env}(t) \tag{6.23d}$$

Ergo, if the power envelope of the AIR is known, the power envelope of the speech can be restored.

Wang and Itakura [384, 386] discuss a multi-microphone approach which attempts to alleviate the effect of the *excess phase* component using a subband envelope technique, assuming that a reference, or training, signal is available for the estimation of the dereverberation filters. Section §6.1.4 discussed the fact that the RTF is usually nonminimum-phase if considered in the full frequency band [274]; however, if the RTF is divided into a number of subbands, only a small proportion of the subbands, considered individually, will have nonminimum-phase characteristics, where the number of bands depends on the bandwidth of each subband and the reverberation time. In general, the RTF between the source and each microphone for a multi-microphone system will be different so, if the number of subbands and microphones is large enough, there will be a high probability that the nonminimum-phase subbands for each RTF will not coincide, and thus a high probability that there will be a least one minimum-phase response for each frequency range. A strategy for selecting the best microphones, so as to select a set of subbands with minimum-phase characteristics, can then be proposed and hence, an estimate of the original speech is possible. The same authors have also investigated a similar approach in single channel dereverberation [385]. Subband multi-microphone techniques are also used since it reduces numerical problems with long impulse responses [385, 403].

Bees *et al*. [26] discuss a single channel blind approach to reverberant speech enhancement using cepstral processing. Cepstral filtering was briefly discussed in section §2.2.4 in the context of signal separation and, as mentioned, signals which are convolved in the time domain have complex cepstra which are additively combined. Furthermore, signals that vary slowly in the cepstral domain can be separated from quickly varying signals by filtering in the cepstral domain. Speech is usually considered as slowly varying in the cepstral domain with its cepstral components concentrated around the cepstral origin, whereas the acoustic impulse response is characterised by pulses with rapid 'ripples' concentrated far away from the cepstral origin,

therefore allowing dereverberation to be achieved by removing the cepstral components corresponding to the impulse response. Segmentation of the speech signal is required for this approach to be successful and relies on a number of heuristics. Single and multi-channel cepstrum-based dereverberation approaches have continued to be an active area of research, for example, see [344, 360].

Cole *et al.* [80] investigate position-independent reverberant speech enhancement using a spectral subtraction method which exploits the 'relatively low effect of position on the spectral magnitude characteristics of the reverberant signal.' A recent attempt at enhancement of reverberant speech is made in Yegnanarayana and Murthy [404] in which the linear prediction (LP) residual signal is identified and manipulated in different regions of the reverberant speech signal, namely, regions in which there is a high signal-to-reverberant component ratio (SRR), low SRR, and reverberant components only. A weighting function is derived to modify the LP residual signal which, in turn, is used to excite a time-varying all-pole filter to obtain perceptually enhanced speech. A rather more direct approach to reverberation cancellation for a particular acoustic environment is the construction of a 'codebook' of all possible *distinct* room transfer functions, from which the appropriate RTF for a particular situation can be estimated, as suggested in [259, 263]. This technique is not ideal for situations where it is not possible to classify all the possible RTFs of a particular acoustic environment.

Further approaches to reverberation enhancement of speech present themselves from a fundamental signal processing approach to signal enhancement:[22]

Direct estimation of the clean speech from the observed reverberant speech by considering the clean speech as 'missing data'. The acoustical properties of the room are considered as unknown, or *nuisance* parameters, and eliminated from the estimation problem through a *marginalisation* process.

Equalisation of the reverberant speech by estimating the room impulse response, and deconvolving the reverberant speech with the *inverse* room response to obtain equalised speech which should approximate the clean speech.

The equalisation of the reverberant speech is an attractive approach since estimation of a large parameter space often leads to instability in the estimation process, as discussed in the next chapter.

[22] A further discussion of these approaches is given in section §7.1.

6.3.2 Excess Phase in Room Transfer Functions

Section §6.1.4 discussed the notion that room transfer functions are usually nonminimum-phase. Since causal inverses of nonminimum-phase systems do not exist, a problem investigated in section §6.3.3, it is important to consider how important the contribution of the nonminimum, or excess, phase in equation (6.16) is to the intelligibility of speech. Johansen and Rubak [169, 170] have considered this question in detail through a number of listening tests:

1. An anechoic speech signal is compared with the same signal filtered by the all-pass component of a real acoustic impulse response.

2. An anechoic speech signal filtered by the minimum-phase component of a real acoustic impulse response is compared with the same anechoic speech signal filtered by the complete impulse response.

The minimum-phase and all-pass components of equation (6.16) are extracted from a nonminimum-phase impulse response using the cepstrum method discussed in section §6.3.3. Results from the first experiment indicate that the all-pass component affects the anechoic signal sufficiently for detection by the human ear. The second experiment indicates that if the excess phase component is removed from the reverberant speech, there is still an audible difference, but less than in the first experiment. This suggests that the minimum-phase component, which contains the magnitude information, is able to partly mask the effect of excess phase. However, in general, the ability of excess phase to degrade speech quality is significant.[23] The work in [169,170] also reinforces the observation that the longer the reverberation time, the more excess phase is present, and the lower the observed speech quality. Moreover, the longer the impulse response and the larger the distance between the source and observation points, the larger the degradation in speech quality. These results are confirmed by considering reverberation distance as discussed in section §6.1.3.4, and the results presented in [257] in which the variation of an acoustic impulse response is investigated for changing source and receiver position and orientations. As the distance between the source and observer positions

[23]The importance of phase in signals has been discussed in depth by Oppenheim and Lim [280], and some of their comments are relevant here, particularly where they argue phase reflects the location of 'events' more than magnitude. Acoustic distortion is mainly due to the arrival, or temporal locations, of 'early' and 'late' reflections. Thus, by the phase argument above, much of this temporal information will be reflected in the phase response rather than the magnitude response and, therefore, it is important to consider excess phase. Note this is in stark contrast to the enhancement of speech in additive white noise, in which it is not as crucial to account for phase distortion [383].

increase, the measure of direct energy to reverberant energy, as discussed in section §6.1.3.2, decreases rapidly, with a rate of decrease higher in rooms with small volume or long reverberation times, than in rooms with a large volume or short reverberation time. As a result, it is crucial that reverberant speech enhancement processes do not neglect the excess phase of the room transfer function. There is further discussion of the importance of phase distortion due to nonminimum-phase systems in a recent paper by Radlović and Kennedy [304].

6.3.3 Invertibility of Room Impulse Response

Reverberant speech enhancement can be achieved by inverse filtering of room acoustics. However, the inversion of a single room transfer function is known to be a difficult task due to the presence of nonminimum-phase zeros which cannot be compensated by causal stable filters [274]. Nevertheless, there are successful solutions for the case of inverse filtering, or equalisation, of room acoustics when there are multiple microphones and either a single source [253, 267, 299], or multiple sources [390]; Haneda *et al*. [145] perform 'multiple-point' equalisation of RTFs using the common-acoustical pole and zero model discussed in section §6.2.6 (also see [105]).[24] Neural networks can also be employed when attempting to equalise nonlinear effects in microphones and loudspeakers [62]. Neely and Allen [274] gave the first comprehensive attempt to design an equaliser for a single linear AIR of a particular room, proposing a minimum-phase equaliser, and this approach is discussed in section §6.3.3.3. The purpose of this section is to discuss two common techniques for inversion of finite length acoustic impulse responses, namely homomorphic [74,140,228,256,279], and least-squares [74,145,253,260]; a comparison of the methods is given in [260].

An important issue in the equalisation of room acoustics is the effect of measurement error of acoustic impulse responses, or variations in acoustics due to changing environmental conditions such as windows or doors opening. It has been shown [80, 257, 305] that if the room is equalised by an inverse filter corresponding to a *different* acoustic response to that which originally degraded the speech, the 'equalised' speech can actually be significantly degraded further. Furthermore, in acoustic echo cancellation (AEC) for 'hands-free' telephones, the acoustic impulse response can change considerably due to the movement of people or objects,

[24] An equivalent problem in, for example, public address systems, is the modification of signals driving loudspeakers, such that the sound heard at a particular point in a room is free of distortion.

and the problem of reverberation removal becomes more complicated. Kerkhof and Kitzen [376] show that the main difficulties in tracking an AIR due to a moving person are due to the large bandwidth of speech, and long reverberation time. Radlović *et al.* [305] quantify errors in equalised AIRs in terms of mean-square deviations, although it is difficult to draw conclusions regarding subjective effects.

6.3.3.1 Problem Formulation

If an acoustic environment has impulse response $\{h(t),\ t \in \mathscr{T}\}$, $\mathscr{T} = \{0, \ldots, T-1\}$, between a source point, $\mathbf{r}_s$, and an observation point, $\mathbf{r}_o$, with a corresponding transfer function, $H(z)$, the aim of deconvolution is to remove the distortions imposed by this filter by designing an *inverse impulse response*, $h_{inv}(t)$, such that:

$$h(t) \star h_{inv}(t) \triangleq \int_{\mathscr{T}} h(\tau)\, h_{inv}(t-\tau)\, d\tau = \delta(t), \quad t \in \mathscr{T} \tag{6.24a}$$

where $\delta(t)$ is the dirac-delta function. The perfect equalisation filter is:

$$H_{inv}(z) = \frac{1}{H(z)} \tag{6.24b}$$

where $h(t) \rightleftharpoons H(z)$ and $h_{inv}(t) \rightleftharpoons H_{inv}(z)$ are $\mathscr{Z}$-transforms. However, the RTF of equation (6.14) is, in general, mixed phase and, as such, does not have a causal and finite inverse given by (6.24b). For reference, Mulgrew [266] proposes nonlinear filters as a solution to the problem of inverting nonminimum-phase systems.

6.3.3.2 Linear Least-Squares Technique

Approximation of inverse filters using the least-squares, or linear predictive, method is well documented and discussed in, for example, [74,145,253,260]. If the least-squares inverse filter is given by $h^{LS}_{inv,\hat{\tau}}(t)$, then $h(t) \star h^{LS}_{inv,\hat{\tau}}(t) = \hat{\delta}(t),\ t \in \mathscr{T}$, where $\hat{\delta}(t)$ is an approximation to the impulse $\delta(t-\hat{\tau})$; *i.e.* $\delta(t)$ delayed by a time $\hat{\tau}$. The time delay $\hat{\tau}$ is incorporated since the inverse of a nonminimum-phase system is, in general, acausal and infinite in length, thus allowing an improved approximation to the inverse of a nonminimum system. If the error between the ideal impulse and the approximated impulse is $e(t) = \hat{\delta}(t) - \delta(t-\hat{\tau})$, then the least-squares approach seeks to minimise the power of the error, $e(t)$, through the cost function:

$$J = \sum_{t \in \mathscr{T}} e^2(t) = \sum_{t \in \mathscr{T}} \left(\sum_{\tau \in \mathscr{T}} h^{LS}_{inv}(\tau)\, h(t-\tau) - \delta(t-\hat{\tau}) \right)^2 \tag{6.25}$$

The filter $h^{LS}_{inv,\hat{\tau}}(t),\ t \in \mathscr{T}$, which minimises J is given by the solution of

$$\mathbf{H}\,\mathbf{h}^{LS}_{inv,\hat{\tau}} = \hat{\mathbf{h}} \tag{6.26a}$$

where the matrix H is given by:

$$[\mathbf{H}]_{t,\tau} = \sum_{\bar{\tau}\in\mathscr{T}} h(\bar{\tau}-\tau)\,h(\bar{\tau}-t), \quad \forall (t,\tau) \in \mathscr{T} = \mathscr{T}\times\mathscr{T} \tag{6.26b}$$

and the vector $[\mathbf{h}^{LS}_{inv,\hat{\tau}}]_t = h^{LS}_{inv,\hat{\tau}}(t),\ t \in \mathscr{T}$. The vector $[\hat{\mathbf{h}}]_t = h(\hat{\tau}-t),\ t \in \mathscr{T}$, simplifies slightly since $h(t)$ is causal, so $h(\hat{\tau}-t) = 0$ if $t > \hat{\tau}$ and, therefore:

$$\hat{\mathbf{h}} = \Big[h(\hat{\tau}),\quad h(\hat{\tau}-1),\quad \ldots,\quad h(0),\quad 0,\quad \ldots,\quad 0\Big] \tag{6.26c}$$

It should be stated that (inverse) filtering is a noise enhancement process, and *white noise gain* is given by $\sum_{t\in\mathscr{T}}[h^{LS}_{inv,\hat{\tau}}(t)]^2$, which will generally be greater than unity.[25] Mourjopoulos *et al.* [260] conclude that zero delay inversion, $\hat{\tau} = 0$, is inferior to optimising the delay, details of the which are also outlined in their paper.

6.3.3.3 Homomorphic Signal Analysis

A second approach to speech dereverberation through inversion of the RTF is *homomorphic deconvolution* [281, Chapter 10], [74, 140, 256, 258–260, 274, 279]. Consider first the decomposition in equation (6.16); Neely and Allen [274] attempt dereverberation by designing an equaliser which equals the inverse of the *equivalent minimum-phase* component:

$$H_{inv}(z) \equiv \frac{1}{H^{PZmp}_{(\mathbf{r}_s,\mathbf{r}_o)}(z)} \tag{6.28}$$

although, for the reasons discussed in section §6.3.2, this approach is not always successful in removing the reverberant component from the impulse response. A number of approaches may be used to extract the minimum-phase component, for instance, Tohyama *et al.* [359] create it by simply taking the absolute value

[25] If WGN $\{w(t),\ t \in \mathscr{T}\}$ with variance σ^2 is passed through a filter $\{g(t),\ t \in \mathscr{T}\}$, then the output is given by the convolution $y(t) = \sum_{\tau\in\mathscr{T}} g(\tau)\,w(t-\tau)$. Therefore,

$$E\left[y^2(t)\right] = \sum_{\tau\in\mathscr{T}}\sum_{\bar{\tau}\in\mathscr{T}} g(\tau)\,g(\hat{\tau})\,E\left[w(t-\tau)\,w(t-\hat{\tau})\right] \tag{6.27}$$

and, since $E\left[w(t-\tau)\,w(t-\hat{\tau})\right] = \sigma^2\,\delta(\hat{\tau}-\tau)$, the noise gain is given by $\sum_{\tau\in\mathscr{T}} g^2(\tau)$.

of the *group delay* of the transfer function. More commonly, the minimum-phase component is extracted using the homomorphic approach on the decompositions in equations (6.15) and (6.16):

$$H^{PZ}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) = H^{PZ\ min}_{(\mathbf{r}_s,\mathbf{r}_0)}(z)\ H^{PZ\ max}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \tag{6.15}$$

$$= H^{PZ\,mp}_{(\mathbf{r}_s,\mathbf{r}_0)}(z)\ H^{PZ\,ap}_{(\mathbf{r}_s,\mathbf{r}_0)}(z) \tag{6.16}$$

The cepstral transform, $\hat{h}(t)$, of the function $h(t)$ is given by:

$$\hat{h}(t) = \mathscr{Z}^{-1}\{\log \mathscr{Z}[h(t)]\} \tag{6.29a}$$

where $\mathscr{Z}$ and $\mathscr{Z}^{-1}$ denote the forward and inverse $\mathscr{Z}$– transforms respectively. The original signal, $h(t)$, can be obtained by performing the inverse operation

$$h(t) = \mathscr{Z}^{-1}\left\{e^{\mathscr{Z}[\hat{h}(t)]}\right\} \tag{6.29b}$$

Taking the cepstral transform of the time-domain equivalent decompositions give the equivalent forms:

$$\hat{h}(t) = \hat{h}_{min}(t) + \hat{h}_{max}(t) \tag{6.30a}$$

$$= \hat{h}_{mp}(t) + \hat{h}_{ap}(t) \tag{6.30b}$$

where the dependence on $(\mathbf{r}_s, \mathbf{r}_0)$ and model choice has been dropped for clarity, and where $\hat{h}(t)$ denotes the cepstral transform of the impulse response $h(t)$.

A complete threatise on homomorphic signal processing is given by Oppenheim and Schafer in their classic book [281, Chapter 10]. Nevertheless, it is instructive to show the origin of some of the properties of the cepstrum transform. The logarithm of the pole-zero model in equation (6.14b) is given by:

$$\ln H(z) = \ln C + \sum_{i=1}^{Q_m} \ln\left[1 - r(i)\, z^{-1}\right] + \sum_{i=1}^{Q_n} \ln\left[1 - s(i)\, z\right] - \sum_{i=1}^{P} \ln\left[1 - p(i)\, z^{-1}\right] \tag{6.31}$$

where the subscripts $(\mathbf{r}_s, \mathbf{r}_0)$ are dropped for convienience and, in order for the complex cepstrum to exist, the time origin of $h(t)$ must be chosen such that $R = 0$. Assuming the $\mathscr{Z}$– transform $H(z)$ exists, the power series expansion for $\ln(1 + x)$

is given by [127]:

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}, \; |x| < 1 \tag{6.32}$$

and, therefore, it is clear that equation (6.31) can be written as:

$$\ln H(z) = -\sum_{k=1}^{\infty} \left[\sum_{i=1}^{Q_n} \frac{s^k(i)}{k} \right] z^k + \ln C + \sum_{k=1}^{\infty} \left[\sum_{i=1}^{P} \frac{p^k(i)}{k} - \sum_{i=1}^{Q_m} \frac{r^k(i)}{k} \right] z^{-k} \tag{6.33}$$

Taking the inverse (noncausal) $\mathscr{Z}$– transform gives:

$$h(t) = \begin{cases} \displaystyle\sum_{i=1}^{Q_n} \frac{s^{-t}(i)}{t} & \text{if } t < 0 \\ \ln C & \text{if } t = 0 \\ \displaystyle\sum_{i=1}^{P} \frac{p^t(i)}{t} - \sum_{i=1}^{Q_m} \frac{r^t(i)}{t} & \text{if } t > 0 \end{cases} \tag{6.34}$$

Therefore, the following properties arise [281, Chapter 10], [279]:

P 1: If $h(t)$ is minimum-phase, then $\{\hat{h}_{min}(t) = \hat{h}_{mp}(t) = 0, \; t < 0\}$

P 2: If $h(t)$ is maximim-phase, then $\{\hat{h}_{max}(t) = \hat{h}_{ap}(t) = 0, \; t > 0\}$

Hence, the complex cepstrum provides a means for factoring a signal $h_{min}(t) \star h_{max}(t)$ into its minimum and maximum-phase components. Specifically, by choosing a linear system operating such that:

$$\hat{h}_1(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{1}{2}\hat{h}(t) & \text{if } t = 0 \\ \hat{h}(t) & \text{if } t > 0 \end{cases} \tag{6.35a}$$

the resulting time-domain signal $h_1(t) = h_{min}(t)$. For the decomposition $h_{mp}(t) \star h_{ap}(t)$, the cepstrum corresponding to the equivalent minimum-phase component is thus:

$$\hat{h}_2(t) = \begin{cases} 0 & \text{if } t < 0 \\ \hat{h}(t) & \text{if } t = 0 \\ \hat{h}(t) + \hat{h}(-t) & \text{if } t > 0 \end{cases} \tag{6.35b}$$

and therefore $h_2(t) = h_{mp}(t)$. Both $h_1(t) \equiv h_{min}(t)$ and $h_2(t) \equiv h_{mp}(t)$ have causal stable inverses, whereas $h_{max}(t)$ has a stable acausal inverse, and $h_{ap}(t)$ has an inverse with an acausal component. The resulting functions $h_1(t)$ and $h_2(t)$ in equation (6.35) can therefore be used for inverse filtering using equation (6.28) or equivalent.

The investigation by Mourjopoulos *et al.* [260] indicates that whilst the least-squares approach has superior equalisation in the time domain, this is not reflected in the frequency domain. Furthermore, it should be noted that the least-squares technique is computationally more efficient than the homomorphic technique.

6.4 Chapter Summary

This chapter has introduced some basic theoretical acoustic properties which are important for understanding why particular models are used throughout this work, and the problems associated with attempting the dereverberation of speech. The suitability of some well-known modelling techniques in signal processing for the representation of room acoustics, the robustness of the models to variations in the source and observer position, and the effect of parameter variation on the accuracy of the model are discussed. Primarily, commonly used models include the *pole-zero*, *all-zero*, *all-pole* and *common-acoustical pole and zero* models. In particular, the *all-zero* model is effective only for a limited spatial combination of source and receiver positions within a particular enclosure [144], and also requires a very large number of coefficients. In contrast, the *all-pole* model is more robust to source–receiver positions and generally requires a lower model order. Finally, some existing approaches for the enhancement of reverberant speech are reviewed, notably the *least-squares* and *homomorphic* approaches. Whilst the least-squares approach has superior equalisation in the time domain to homomorphic techniques, this is not reflected in the frequency domain. The issue of the contribution of nonminimum-phase to the perception of reverberation has been discussed, and it is important that this component is not neglected since it contains most of the reverberant energy.

7

Introduction to Blind Deconvolution

THE remaining chapters of Part III consider the *blind deconvolution* problem, with application to reverberation cancellation of speech signals in acoustic environments. Many of the blind deconvolution techniques applied to adaptive equalisation of communication channels assume that the source signal is contained within a finite support [152] and, or, are independent and identically distributed (i. i. d.) [57, 70]. This *prior* knowledge dramatically changes the type of estimation problem. Blind deconvolution of discrete input linear systems has been considered in a non-Bayesian framework [109, 215–218], and within a Bayesian framework [70], while blind deconvolution of general systems with i. i. d. inputs has been considered in Cappé *et al*. [57]. Bayesian blind deconvolution of point, or impulsive processes, is considered in Andrieu *et al*. [12]. However, for dereverberation of continuous speech signals, in which the source signal is highly correlated and belongs to a set of infinite support, the techniques applied to many adaptive equalisation algorithms cannot be applied directly. In fact, to use these techniques, the source signal (with infinite support) must be decomposed into a linear time-varying filter excited by an i. i. d. process.

Moreover, the blind deconvolution techniques that assume a finite support for

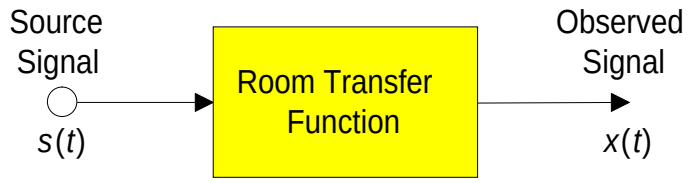


Figure 7.1: *A signal model for single channel blind deconvolution.*

the source signal often assume *short term stationarity* of the system, and do not take *global nonstationarity* into account. As discussed in Chapter 1, utilising the global nonstationarity of a system allows the identification of system characteristics which may otherwise be unobservable; this is demonstrated in the case of the separation of nonstationary signals, as outlined in Part II. There must, however, be some distinguishing features between the speech signal and the room acoustics to allow blind deconvolution of the observed signal. Note, then, that speech possesses *nonstationary* statistical characteristics, while the acoustic properties of a room essentially define a *stationary* system, *provided* that the sound source and observer are spatially stationary. Although both signals belong to a set with infinite support, this distinguishing feature, if utilised, provides enough information which leads to a new technique for blind deconvolution as discussed in the following chapters. The general signal model for the blind deconvolution problem is shown in Figure 1.6, and the problem is formally defined below.

7.1 Blind Deconvolution Problem Statement

In the single channel blind deconvolution problem, a degraded observed signal,[1] $\{x(t)\} \in \mathbb{R}^T$, $t \in \mathscr{T} = \{1, \ldots, T\} \subset \mathbb{Z}$, is modelled as the convolution of the source signal, $\{s(t)\} \in \mathbb{R}^T$, with a distortion operator, $\mathscr{A}$. The distortion operator could, for example, represent the acoustical properties of a room, as shown in Figure 7.1, the effect of *multipath propagation* in the reception of radio signals, or a non-impulsive excitation in seismic applications. The case when the source signal is stationary, and $\mathscr{A}$ is a nonlinear function, has been considered by Troughton and God-

[1] Note in this part of the dissertation, the Bayesian methodology is employed and, therefore, *all* variables are considered to be random. Thus, bold symbols exclusively denote vector or matrices, with the former always in lower case, and the latter in upper case.

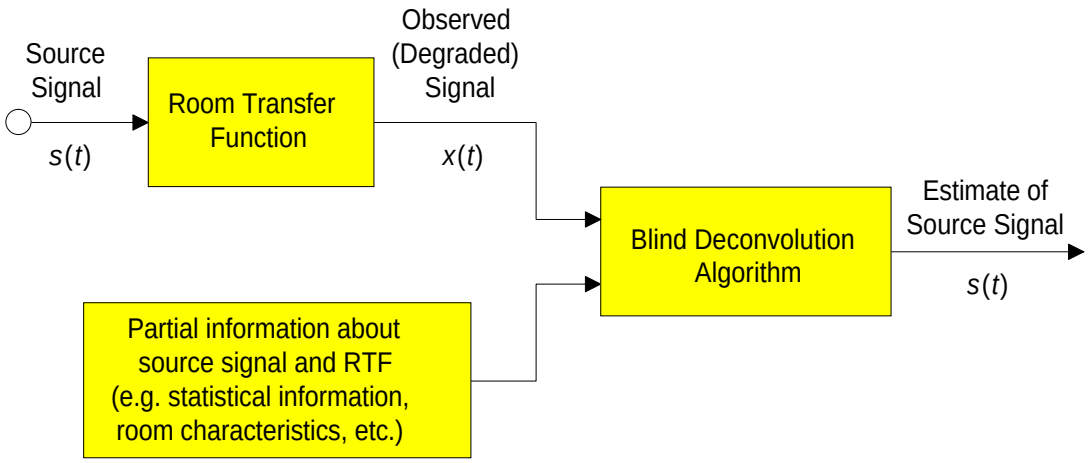


Figure 7.2: *General blind deconvolution scenario.*

sill [363,364].[2] Since the acoustic response of a typical room is, in general, a combination of time delays and linear attenuation [8,59,111,143,168,191,201,262,354] (also see Chapter 6), it fits a linear model with a reasonable degree of accuracy and, as such, only linear distortion operators are considered in this thesis. Hence, room responses can be represented by the *impulse response* function, $h(t,\tau)$, with the source and observed signals related by:

$$x(t) = h(t,\tau) \star s(\tau) \tag{7.1}$$

where $\star$ denotes convolution over the variable τ as introduced in equation (3.35). The task is to estimate the source signal, $s(t)$, *or* the distortion operator, $\mathscr{A}$, given only the observed signal, $x(t)$. The problem is under-constrained and *can only* be solved by incorporating varying degrees of *a priori* knowledge regarding $\{s(t)\}$ and $\mathscr{A}$. The general distinguishing features of the signal $s(t)$ and the operator $\mathscr{A}$ *must* be known, otherwise it becomes impossible to assess when the algorithm is performing correctly. The general blind deconvolution scenario is shown in Figure 7.2 [198]. Several important characteristics of blind deconvolution include:

1. The source signal and impulse response of the distortion operator must be *irreducible* for unambiguous deconvolution [198]. An (real) irreducible signal is one that cannot be expressed as the convolution of two or more signal components, on the understanding that the delta function is not a sig-

[2] A common example of nonlinear distortion is in the quantisation of analogue signals for representation in digital form [363, Chapter 7].

nal component. Suppose the distortion operator, $\mathscr{A}$, is LTI, then the observed signal may be expressed as $x(t) = h(t) \star s(t)$. If either $h(t)$ or $s(t)$ are *reducible* such that $h(t) = h_1(t) \star h_2(t)$, and $s(t) = s_1(t) \star s_2(t)$, then $x(t) = h_1(t) \star h_2(t) \star s_1(t) \star s_2(t)$. It is impossible to decide which component belongs to the source signal or distortion operator without additional knowledge. It is shown that the use of nonstationarity can compensate for the case where $s(t)$ or $\mathscr{A}$ are, in fact, *reducible* over a short period of time.

2. In general the goal in blind deconvolution is to obtain a scaled shifted version of the original signal, $\hat{s}(t) = a\, s(t - \tau)$, where $a, \tau \in \mathbb{R}$ are constants.

There are two distinct approaches to estimating $s(t)$:

1. Estimate directly $s(t)$, or the parameters and excitation of an appropriate parametric model, as a 'missing data' problem by treating the parameters of $\mathscr{A}$ as *nuisance parameters*, or,

2. Model $\mathscr{A}$ by a LTV filter, estimate the filter coefficients by treating $\{s(t)\}$ as a nuisance parameter, and then deconvolve $x(t)$ with $\mathscr{A}$ to recover $s(t)$.

For the reasons discussed in the following sections, the latter approach is considered in this dissertation. The characteristics of the blind deconvolution problem introduced here can be extended to the two-dimensional case for the restoration of blurred images, and in section §11.2.1 this is proposed as future work.

7.2 Homomorphic Blind Deconvolution

An excellent review of blind deconvolution and blind equalisation for communication systems may be found in Haykin [152]. Many of the techniques in the literature deal with the problem of multichannel blind deconvolution [119], or assume considerable knowledge of the source signal, as discussed in this chapter's introduction. This section discusses the significant contribution of Stockham *et al.* [337] called the *homomorphic* approach to single channel blind deconvolution, for the restoration of images and old recordings.

This homomorphic technique assumes that another 'clean' signal, spectrally similar to the desired source signal, is available for use as a *nonparametric spectral template*. The first stage of the process is to transform the convolution process

into one of addition so that conventional filtering techniques may be used on the resulting signal. Therefore, since the observed signal $x(t) = h(t) \star s(t)$, $t \in \mathscr{T}$, where $h(t)$ is the impulse response of the distortion operator $\mathscr{A}$, the logarithm of the Fourier transform of $x(t)$ is given by:

$$\log X(e^{j\omega}) = \log H(e^{j\omega}) + \log S(e^{j\omega})$$

Suppose that the observed signal $\{x(t),\, t \in \mathscr{T}\}$ is divided into M intervals, or *blocks*, each block, i, containing samples $\{x_i(t),\, t \in \{0, \ldots, T_i - 1\}\} \equiv \{x(t),\, t \in \{t_i, \ldots, t_{i+1}\}$, where $T_i = t_{i+1} - t_i$, $i \in \mathscr{M} = \{1, \ldots, M\}$. In each block, the distortion operator, $\mathscr{A}$, is assumed to have the same characteristics, but the spectrum of the source signal is assumed to be, in general, different, although this isn't a requirement. If the block length, T_i, is large compared with the temporal extent of the impulse response, $h(t)$, windowing effects are small and the approximate relationship:

$$X_i(e^{j\omega}) \approx H(e^{j\omega})\, S_i(e^{j\omega}) \tag{7.2}$$

results, where $S_i(e^{j\omega})$ is the Fourier transform of the source signal samples contained only in block i. Taking complex logarithms[3] and averaging over all the data blocks gives, after equating real and imaginary parts;

$$\frac{1}{M}\sum_{i=1}^{M} \log\left|X_i(e^{j\omega})\right| \approx \frac{1}{M}\sum_{i=1}^{M} \log\left|S_i(e^{j\omega})\right| + \log\left|H(e^{j\omega})\right| \tag{7.3a}$$

$$\frac{1}{M}\sum_{i=1}^{M} \angle X_i(e^{j\omega}) \approx \frac{1}{M}\sum_{i=1}^{M} \angle S_i(e^{j\omega}) + \angle H(e^{j\omega}) \tag{7.3b}$$

Since the complex logarithm is multivalued, there are difficulties in calculating the phase of the spectra in such a manner to be compatible with the phase equation (7.3b). Therefore, assuming the ear is relatively insensitive to phase properties, Hilbert transform techniques are used to obtain a minimum-phase response for use in equation (7.3b) [281]. The important step is to assume that a clean spectral template is available from, for example, a modern recording where the frequency response of the recording equipment is flat. Then, assuming $\{S'_k(e^{j\omega})\}$, $k \in \mathscr{K} = \{1, \ldots, K\}$, are spectral estimates of the clean signal over K intervals, perhaps *different* to those chosen for the distorted recording, the following approximation is

[3] $\log z = \log(|z|\ e^{j\angle z}) = \log|z| + j\angle z$

made:

$$\frac{1}{M}\sum_{i=1}^{M}\log\left|S_i(e^{j\omega})\right| \approx \frac{1}{K}\sum_{k=1}^{K}\log\left|S_i'(e^{j\omega})\right| \tag{7.4}$$

Hence, substituting equation (7.4) into equation (7.3a) leads to an estimate for $H(e^{j\omega})$ as the problem has now been reduced to one of deconvolution. It is reported that restoration of archived records in good condition and without surface noise, retains 'acoustic flavour', removes prominent surges in volume when the pitch of the voice or music strikes the recording equipment resonances,[4] and eliminates the megaphone effect; overall the restoration is reported as 'very pleasing'.

Stockham *et al*. [337] goes further to discuss the effect of noise, and presents an elegant result using a similar process to that outlined above. However, this approach relies on the rather poor assumption that the source signal as stationary. Further, a *power spectral* technique is discussed, in which averaging is performed directly on the squared magnitude of the Fourier transform of the data segments, and it is shown that the estimator yields the same restoration filter as the homomorphic approach, but possessing a lower variance.

This homomorphic approach relies on two important factors:

- First, the distortion operator is assumed stationary and, therefore, its characteristics remain the same over all data blocks.

- Second, a significant amount of prior knowledge regarding the source signal is required.

The issue of prior knowledge and *degrees of nonstationarity* are once again apparent in successful blind deconvolution techniques. The approach developed in this work aims to remove the requirement of finding a suitable spectral envelope to match that of the source signal and, therefore, requires considerably less prior knowledge.

[4]Generally, older recordings were made through an acoustical horn.

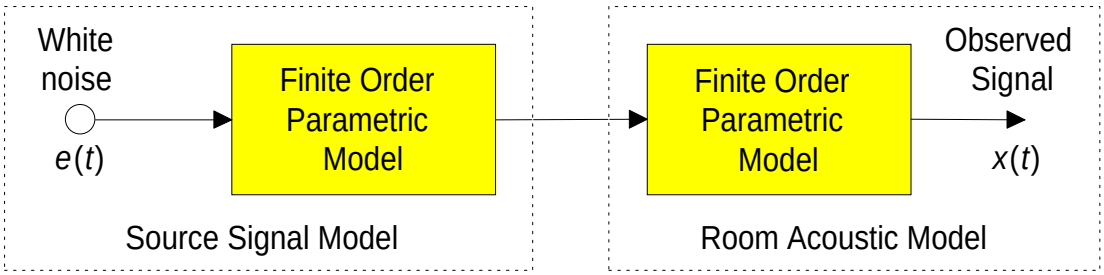


Figure 7.3: *General parametric model of the blind deconvolution problem.*

7.3 System Modelling

The theory and results presented in Part II demonstrates that, in order to perform signal separation for general nonstationary signals, some form of signal structure must be known *a priori*; section §4.6, for example, discussed separation for the case when the signals were known to fit the *uniform modulated* model. This is also true for blind deconvolution, in which some prior knowledge regarding the structure of the source signal and channel must be made available to recover the source signal. Therefore, the second approach to blind deconvolution discussed in section §7.1 is considered, and various models will be used for the source signal, $\{s(t)\}$, and the distortion operator, $\mathscr{A}$. These models are assumed to be *causal*, *linear*, *time-varying* (where appropriate), and of finite-order. The resulting general finite-order parametric model for the system of Figure 7.1 is shown in Figure 7.3, and the models which lead to analytically tractable results will be discussed in the following sections. There are, in fact, two distinct modelling techniques which are used in the modelling of the deconvolution problem, namely *input–output modelling*, and *time series modelling*

Input–output modelling seeks to model the *transfer function* relating the input of a system to its output. The acoustic properties of a room is one such example, relating the sound source to the observed sound in a room (see section §3.3 for a further discussion on systems modelled by a LTV process).

Time series modelling seeks to model a *stochastic process* in terms of the output of a (LTV) system excited by an orthogonal stationary stochastic process (see section §3.2.3.2).

Hence, $\mathscr{A}$ is modelled using the former technique, whilst $s(t)$ is modelled by the latter. The most common input–output models are considered in section §7.4, and common time series models are considered in section §7.5. The inter-relationships between these models are also discussed throughout these sections.

7.4 Linear Input–Output Modelling

In section §2.1.1 of Part II, it was observed that currently no general theory exists to deal with nonlinear filters. Although there are many nonlinear modelling techniques, they are not easily categorised and, in general, a different analytical technique must be used for each model. As discussed in section §7.1, the acoustic properties of a typical room fit linear models with a reasonable degree of accuracy and, therefore, the acoustic models discussed in this dissertation are restricted to linear finite-order parametric input–output models. These linear models may be either *linear time-invariant* (LTI), or *linear time-varying* (LTV). The transfer function due to the acoustics of a room generally do not change considerably with time, but only with the spatial locations of the sound source and observer. Therefore, if the source and observer are spatially stationary, a LTI model is appropriate, or a LTV model if the source or the observer are moving.

7.4.1 Time-Invariant Linear Model

The general finite-order *linear time-invariant* parametric system model relates the input of the system $\{s(t),\ t \in \mathscr{T}\}$ to the output of the system $\{x(t),\ t \in \mathscr{T}\}$ by:

$$x(t) = \underbrace{-\sum_{p=1}^{P} a(p)\, x(t-p)}_{\text{All-pole model}} + \underbrace{\sum_{k=0}^{K} g(k)\, s(t-k)}_{\text{All-zero model}} \tag{7.5}$$

where $\{a(p),\ p \in \mathscr{P} = \{1, \ldots, P\}\}$ and $\{g(k),\ k \in \mathscr{K} = \{1, \ldots, K\}; g(0) \triangleq 1\}$ are the model parameters. This model has P all-pole terms, and K all-zero terms.

7.4.2 Time-Varying Linear Model

The general finite-order *linear time-varying* parametric system model relates the input of the system $\{s(t),\ t \in \mathscr{T}\}$ to the output of the system $\{x(t),\ t \in \mathscr{T}\}$ by:

$$x(t) = \underbrace{-\sum_{p=1}^{P_t} a(t,p)\, x(t-p)}_{\text{All-pole model}} + \underbrace{\sum_{k=0}^{K_t} g(t,k)\, s(t-k)}_{\text{All-zero model}} \tag{7.6}$$

where $\{a(t,p),\ p \in \mathscr{P}_t = \{1,\ldots,P_t\},\ t \in \mathscr{T}\}$ and $\{g(t,k),\ k \in \mathscr{K}_t = \{1,\ldots,K_t\}; g(t,0) \triangleq 1; t \in \mathscr{T}\}$ are the time-varying parameters which may either be modelled as a deterministic time-varying function, or as an unobserved random process. At each time instance, this model has P_t all-pole terms, and K_t all-zero terms and, therefore, the system is represented by $P_t + K_t$ parameters. Clearly, this is an over-parameterised model and, therefore, degenerate. Given a *single* realisation of the input–output relationships, it is impossible to uniquely estimate the model parameters. This issue is discussed further in section §7.5.2.1.

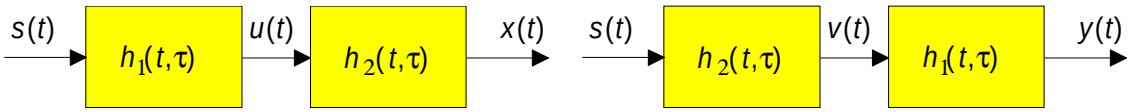
7.4.3 Poles and Zeros of a Linear Time-Varying System

Poles and zeros are well defined notions in linear time-varying system theory and are useful for characterising a particular system as well as providing information about the stability of a system. A linear discrete-time system, given by the difference equation of (7.5), may be expressed in terms of the system poles $\{p(i),\ i \in \mathscr{P}\}$ and zeros $\{q(k),\ k \in \mathscr{K}\}$ by:

$$\left[\prod_{i=1}^{P} \left(1 - p(i)\, z^{-1}\right)\right] x(t) = \left[\prod_{k=1}^{K} \left(1 - q(k)\, z^{-1}\right)\right] s(t), \quad t \in \mathscr{T} \tag{7.7}$$

where z denotes the left-shift operator defined as $z^{-i}\, x(t) \triangleq x(t-i)$, where $t \in \mathbb{Z}, i \in \mathbb{Z}$. LTV systems defined by the time-varying difference equation in (7.6), can be written in the form,

$$A(z,t)\, x(t) = G(z,t)\, s(t) \tag{7.8a}$$



(a) System 1 preceding system 2.

(b) System 2 preceding system 1.

Figure 7.4: *Series connection of two linear time-varying (LTV) systems with impulse responses* $h_i(t,\tau)$, $i \in \{1,2\}$.

where the time-dependent polynomials $A(z,t)$ and $G(z,t)$ are given by,[5]

$$A(z,t) = 1 + \sum_{i=1}^{P} a(t,i)\, z^{-i} \quad \text{and} \quad G(z,t) = \sum_{k=0}^{K} g(t,k)\, z^{-k} \tag{7.8b}$$

There are two notions for the poles and zeros of the system:[6] the so-called *frozen-state approximation* [417] and the time-dependent pole factorisation discussed by Kamen [175]. The need for these two definitions arise from the observation in section §3.3.2.2 that the cascade series connection of two linear time-varying systems with impulse responses $h_1(t,\tau)$ and $h_2(t,\tau)$ is noncommutative. This can be seen by considering the two systems shown in Figure 7.4; if the impulse response of the system in Figure 7.4(a) is $h_{12}(t,\tau)$ and that of Figure 7.4(b) is $h_{21}(t,\tau)$ then, using the superposition integral (3.35),

$$h_{12}(t,\tau) = \int_{\mathscr{T}} h_1(\hat{\tau},\tau)\, h_2(t,\hat{\tau})\, d\hat{\tau} \quad \text{and} \quad h_{21}(t,\tau) = \int_{\mathscr{T}} h_1(t,\hat{\tau})\, h_2(\hat{\tau},\tau)\, d\hat{\tau} \tag{7.9}$$

which clearly, in general, are not identical. The notion of poles and zeros for a LTV system is developed in the following two sections.

7.4.3.1 Poles of a Second-Order LTV System

Consider an intuitive implementation of, for example, a second-order time-varying difference equation in terms of some 'system poles' $p_1(k)$ and $p_2(k)$ as shown in

[5] $A(z,t),\ G(z,t) \in \Gamma(z)$, where $\Gamma(z)$ is the set of all polynomial operators with time-dependent parameters [264]:

$$\Gamma(z) \triangleq \left\{ \sum_{i=0}^{\infty} f(t,i)\, z^{-i},\ \forall f(\cdot,i) \in \Upsilon \triangleq \{f : \mathbb{Z} \to \mathbb{R}\},\ \forall i \in \mathbb{Z},\ f(t,0) \neq 0,\ \forall t \in T \subset \mathbb{Z} \right\}$$

[6] Note that the number of poles and zeros are assumed to be fixed in this definition.

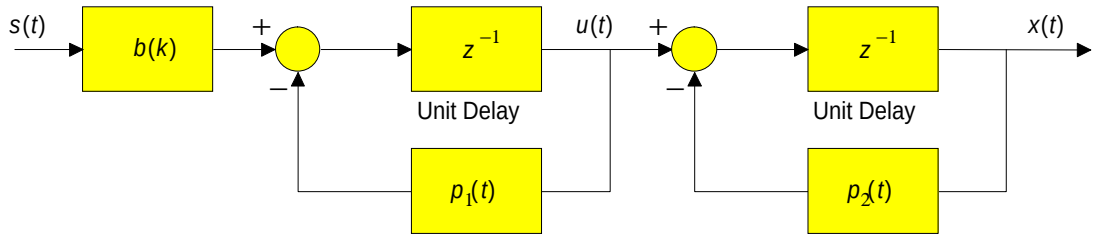


Figure 7.5: *Realisation of second-order time-varying difference equation.*

Figure 7.5. The system poles $p_1(t)$ and $p_2(t)$ are so-called by considering the first stage of the system where its output $u(t)$ is given by:

$$[z - p_1(t)]\, u(t) = b(t)\, s(t) \tag{7.10}$$

Comparing the form of (7.10) with the equivalent form for a LTI system, a natural definition for the system poles is $p = p_1(t)$. Given this definition, the output of the system is given by:

$$\begin{aligned} \{[z - p_1(t)] \circ [z - p_2(t)]\}\, x(t) &= b(t)\, s(t) \\ \left[z^2 - (p_1(t) + p_2(t+1))\, z + p_1(t)\, p_2(t)\right]\, x(t) &= b(t)\, s(t) \end{aligned} \tag{7.11}$$

where $\circ$ denotes skew multiplication defined by the relationship in equation (7.11).[7] Comparing this expression with the form in equations (7.8):

$$\left[z^2 + a_0(t)\, z + a_1(t)\right]\, y(t) = b(t)\, s(t) \tag{7.12}$$

then $p_1(t)$ and $p_2(t)$ must satisfy:

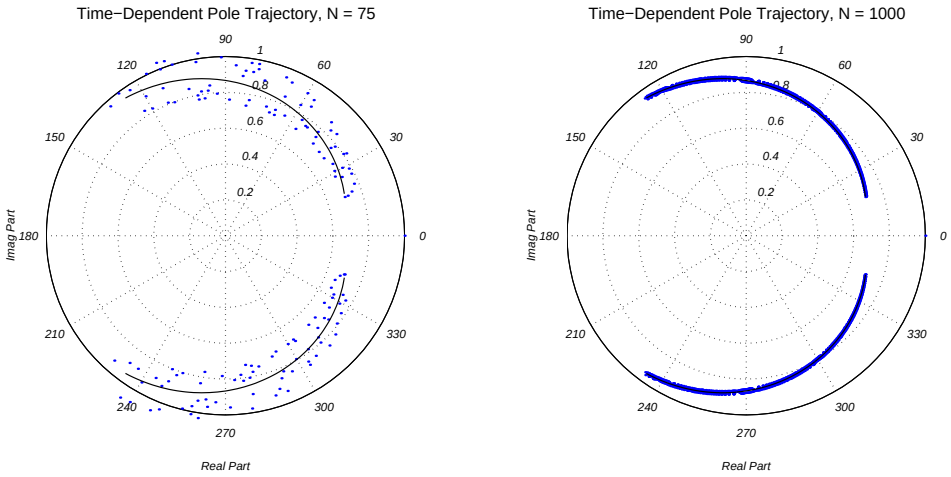
$$p_1(t) + p_2(t+1) = -a_0(t) \tag{7.13a}$$

$$p_1(t)\, p_2(t) = a_1(t) \tag{7.13b}$$

which may be combined into a single nonlinear first-order difference equation in terms of $p_2(t)$:

$$p_2(t+1)\, p_2(t) + a_0(t)\, p_2(t) + a_1(t) = 0 \tag{7.13c}$$

[7] Also see footnote 8 on page 164 in section §7.4.3.2.



(a) Rapid-varying system.

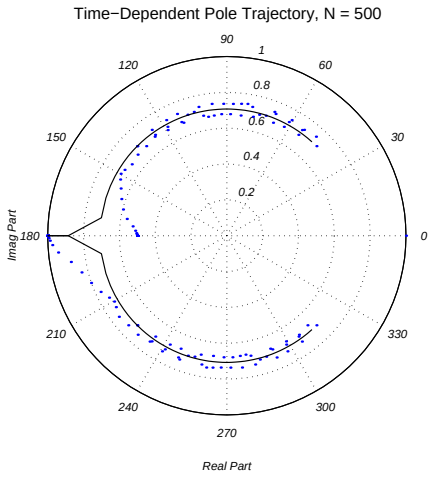
(b) Slowly-varying system.

Figure 7.6: *Comparison of time-dependent (dots) and frozen-state (solid line) pole trajectories for a second-order difference equation.*

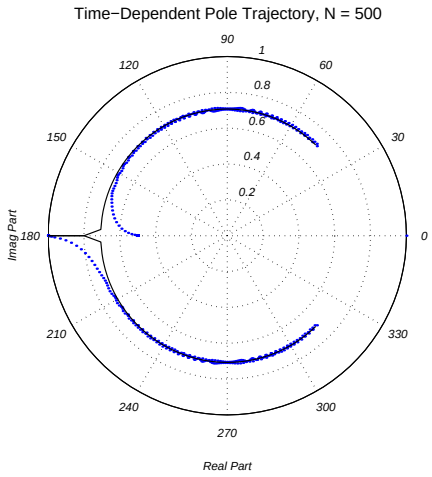
Such a definition produces a homogeneous solution for the output of the LTV system analogous to those obtained in the LTI case [175]. If a system is *slowly* varying, in the sense that the distance between the time-varying pole locations $p_2(t)$ at consecutive sample instances is small, then it can be studied using the frozen-state approximation where the time-varying poles are defined as the roots of the polynomial in equation (7.12): $\left[z^2 + a_0(t)\, z + a_1(t)\right]$, where t is considered as a parameter. As an example, there is a marked difference between *Kamen's time-dependent poles* defined by equations (7.13), and the frozen-state approximation, as shown in Figure 7.6 for the case when,

$$\left.\begin{aligned} a_0(t) &= -2r(t)\cos\theta(t) \\ a_1(t) &= r^2(t) \end{aligned}\right\} \text{ where } \begin{cases} r(t) = r_{\min} + \dfrac{r_{\max} - r_{\min}}{T}\, t \\ \theta(t) = \theta_{\min} + \dfrac{\theta_{\max} - \theta_{\min}}{T}\, t \end{cases} \tag{7.14}$$

and $t \in \mathscr{T} = \{1, \ldots, T\} \subset \mathbb{Z}$, where T is the number of samples and $\{r_{\min},\, \theta_{\min}\}$, $\{r_{\max},\, \theta_{\max}\}$ determine the initial and final locations of the frozen-state time-varying pole. The frozen-state poles are located at $p_{fs}(t) = r(t)\, e^{\pm j\theta(t)}$, $t \in \mathscr{T}$. The difference in magnitude between Kamen's time-dependent poles and the frozen state approximation is noticeable for rapidly changing systems as shown in Figure 7.6(a) and becomes negligible when the system is slowly changing – see Figure 7.6(b).



(c) Rapid-varying system.



(d) Slowly-varying system.

Figure 7.6: *Continued.*

Note that if $p_1(t)$ and $p_2(t)$ are the solutions to equations (7.13) with initial value $p_2(t_0) = \gamma \in \mathbb{C}$, $t_0 \in \mathscr{T} \subset \mathbb{Z}$, then the complex conjugates $\bar{p}_1(t)$ and $\bar{p}_2(t)$ are also solutions to (7.13) with $p_2(t_0) = \bar{\gamma}$. The elements $p_2(t)$ and $\bar{p}_2(t)$ are called the *right poles* for $t > t_0$, and $p_1(t)$ and $\bar{p}_1(t)$ are the *left poles* for $t > t_0$. Unlike the time-invariant case, in general, $\bar{p}_1(t) \neq p_2(t)$ and $\bar{p}_2(t) \neq p_1(t)$ and, as such, there are generally two different solutions to (7.13) corresponding to the fact that the cascade series connection of two LTV systems is noncommutative. Kamen [175] shows that, in fact,

$$p_1(t) = \left[\frac{\bar{p}_2(t+1) - p_2(t+1)}{\bar{p}_2(t) - p_2(t)}\right] \bar{p}_2(t) \tag{7.15}$$

and for slowly varying systems the term in the square brackets is approximately unity and, therefore, $p_1(t) \to \bar{p}_2(t)$. This is vividly demonstrated in the example discussed in [175] with coefficients:

$$a_0(t) = 0.5, \text{ for all } t \in \mathbb{Z}, \quad \text{and} \quad a_1(t) = \begin{cases} -1, & k \leq 0, \\ -1 + \frac{2.5t}{T}, & 0 < t < T, \\ 1.5, & t \geq T \end{cases} \tag{7.16}$$

Figures 7.6(c) and 7.6(d) show the pole trajectories for the rapidly and slowly vary-

ing case described by (7.16). The erratic behaviour of the pole in Figure 7.6(c) as t increases from its initial value is a result of the rapid variation of the coefficient $a_1(t)$. Kamen [175] discusses the second-order case in further detail.

7.4.3.2 Higher-Order LTV Systems

The results from the previous section may be generalised by writing (7.8) as:

$$\left[\prod_{i=1}^{P} \circ \left(1 - p_i(t)\, z^{-1}\right)\right] x(t) = \left[\prod_{k=1}^{K} \circ \left(1 - q_k(t)\, z^{-1}\right)\right] s(t) \tag{7.17}$$

where $\prod \circ$ denotes skew multiplication.[8] As such, the polynomial operators defined in equations (7.8) can be factorised in terms of the functions $\{p(t,i)\}_{i=1}^{P}$ and $\{q(t,k)\}_{k=1}^{K}$, referred to as the model *poles* and *zeros* respectively, such that [175],

$$\begin{aligned} A(z,t) &\triangleq \left[1 - p(t,1)\, z^{-1}\right] \circ \left[1 - p(t,2)\, z^{-1}\right] \circ \cdots \circ \left[1 - p(t,P)\, z^{-1}\right] \\ G(z,t) &\triangleq \left[1 - q(t,1)\, z^{-1}\right] \circ \left[1 - q(t,2)\, z^{-1}\right] \circ \cdots \circ \left[1 - p(t,P)\, z^{-1}\right] \end{aligned} \tag{7.18}$$

where $p(t,1)$, $\{p(t,i)\}_{i=2}^{P-1}$ and $p(t,P)$ are respectively called the *left*, *inner* and *right* poles of the all-pole model; corresponding terminology is used for the model zeros. Assuming the polynomials $A(z,t)$ and $G(z,t)$ are known, Kamen's time-dependent poles can be calculated using the algorithm presented in section §7.4.3.3. Again, if a system is *slowly* varying, it can be studied using the frozen-state approximation where the time-varying poles and zeros, given by $\{p_{fs}(t,i),\, i \in \mathscr{P}\}$ and $\{q_{fs}(t,i),\, i \in \mathscr{K}\}$ respectively, are defined as the roots are the polynomials $A(z,t)$ and $G(z,t)$ respectively, where t is considered as a parameter. Further discussion of the poles and zeros of LTV systems can be found in [175, 264, 271].

7.4.3.3 Factorisation of Kamen's Time-Dependent Poles

The poles of equation (7.17) can be estimated from the time-varying coefficients $a(t,p)$ by noting they satisfy the following equation:

$$\left[1 + \sum_{i=1}^{P} a_P(t,i)\, z^{-i}\right] x(t) \triangleq A_P(z,t)\, x(t) = \left[\prod_{i=1}^{P} \circ \left(1 - p_i(t)\, z^{-1}\right)\right] x(t) = 0 \tag{7.19}$$

[8] Skew multiplication is noncommutative and is defined such that $z^{-i} \circ z^{-k} = z^{-(i+k)}$, and $z^{i} \circ a(t) = a(t-i)\, z^{-i}$.

where the $a(t,i)$'s have been relabelled $a_P(t,i)$ to facilitate the development of a recursive algorithm. Assuming the existence of the *right* pole $p(t,P)$, (7.19) may be written:

$$\left\{\left[1+\sum_{i=1}^{P-1} a_{P-1}(t,i)\,z^{-i}\right]\circ\left[1-p(t,P)\,z^{-1}\right]\right\}x(t) \\ \triangleq \left\{A_{P-1}(z,t)\circ\left[1-p(t,P)\,z^{-1}\right]\right\}x(t)=0$$

Expanding, using the definition of the skew operator, yields a polynomial in the same form of (7.19) and, by comparing coefficients in z^{-i}, it follows [175, 271]:

$$a_P(t,i)=\begin{cases} a_{P-1}(t,1)-p(t,P) & \text{if } i=1,\\ a_{P-1}(t,i)-a_{P-1}(t,i-1)\,p(t-i+1,P) & \text{if } 2\le i\le P-1,\\ -a_{P-1}(t,P-1)\,p(t-P+1,P) & \text{if } i=P \end{cases} \tag{7.20a}$$

which may, alternatively, be written as:

$$a_{P-1}(t,i-1)=\begin{cases} -\dfrac{a_P(t,P)}{p(t-P+1,P)} & \text{for } i=P,\\ \dfrac{a_{P-1}(t,i)-a_P(t,i)}{p(t-i+1,P)} & \text{for } i=\{P-1,\ldots,2\}, \end{cases} \tag{7.20b}$$

$$p(t,T)=a_{P-1}(t,1)-a_P(t,1)$$

Hence, given the initial pole positions $\{p(\tau,P),\ \tau=\{t-P+1,\ldots,t-1\}\}$, it is possible to calculate $a_{P-1}(\tau,i)$, $i\in\{1,P\}$ and, therefore, $p(\tau,P)$, $\tau\ge t$. The polynomial $A_{P-1}(z,t)$ can then be factorised as $A_{P-2}(z,t)\circ\left[1-p(t,P-1)\,z^{-1}\right]$, and so forth, such that all the poles are calculated. A similar procedure can be applied to calculate the zeros.

7.5 Linear Time Series Modelling

A stochastic process can be modelled as the output of a LTV filter which is excited by a zero mean white noise process. Therefore, the time-series models discussed below have the same form as the linear filter models discussed in section §7.4. However, since the time-series are stochastic processes, the terminology is slightly

different, despite the fact that the models have essentially the same form. There are two distinct classes of linear time-series models: *stationary* and *nonstationary*. A further class is also considered for signals whose statistics vary slowly with time. In this case the signal may be considered *locally* stationary over a short period of time, but *globally* nonstationary. A process with this property is called *quasi-stationary* or, sometimes, *short-term stationary*. A quasi-stationary signal may be modelled using a *stationary* time-series model over a given time segment; for example, a typical audio signal can be considered quasi-stationary over periods of around 25 ms. Quasi-stationary models are discussed in section §7.5.3.

7.5.1 Stationary Models

Stationary models possess stationary statistics and, as such, are usually applied to model stationary signals. Widely used models include the *autoregressive* (AR) model, which is effective for modelling speech and audio signals [18,221,237], the *moving average* (MA) model, which is a weighted average of previous inputs [353], and the autoregressive moving average (ARMA) model, which is a general case of the AR and MA models [353]. The MA model is useful when the predominant features of a signal's spectrum are nulls at specific frequencies. The AR model is used when the spectral features include dominant peaks, and the ARMA model can represent both nulls and peaks.

7.5.1.1 ARMA Model

The *autoregressive moving average* model is the most general linear time-series model, of which the AR and MA models are special cases. This model incorporates both an autoregressive term, which depends on the previous values of the time series being modelled, and a moving average term, which depends on previous excitation samples. A time series $\{s(t)\}$ can be expressed in terms of an excitation sequence $\{e(t)\}$, which is assumed to be a *stationary* zero mean white Gaussian noise (WGN) process, $e(t) \sim p_E\left(e(t) \,\middle|\, \sigma_e^2\right) = \mathcal{N}\left(e(t) \,\middle|\, 0,\, \sigma_e^2\right)$, as:

$$s(t) = \underbrace{-\sum_{q=1}^{Q} b(q)\, s(t-q)}_{\text{AR terms}} + \underbrace{\sum_{r=0}^{R} c(r)\, e(t-r)}_{\text{MA terms}} \tag{7.21}$$

where $\{b(q),\ q \in \mathscr{Q} = \{1, \ldots, Q\}\}$ and $\{c(r),\ r \in \mathscr{R} = \{1, \ldots, R\}; c(0) \triangleq 1\}$ are the model parameters. This model has Q AR terms, R MA terms, and is denoted a ARMA(Q, R) model. Note that this model is equivalent to passing the excitation sequence, $e(t)$, through the LTI filter described by equation (7.5).

7.5.1.2 AR Model

The autoregressive model is a very popular time series modelling approach, mainly due to the fact that accurate estimates of the AR parameters can be obtained by solving a set of linear equations, and also because the spectral characteristics of the AR model resembles the resonances of the vocal tract. As such, AR models have been extensively used in speech processing [117, 120, 130, 221, 237, 378, 379, 381]. The ARMA model in equation (7.21) for the case when $R = 0$ reduces to the AR(Q) model given by:

$$s(t) = -\sum_{q=1}^{Q} b(q)\, s(t-q) + e(t) \tag{7.22}$$

The statistical properties for the autoregressive model and *stochastic difference equations* have been studied extensively in the literature [49, 238], and much is known about the statistical properties of various autoregressive spectral estimators [184–186, 205, 286, 326]. The problem of determining the model order has received much attention in the literature since overestimation of the model order can cause erroneous peaks in the estimated spectrum and high variance for the model parameters, while underestimation may lead to a smoothed spectrum. For model order determination, there exists a long list of approaches ranging from *information criteria* [53, 92, 135, 388] to *Bayesian methods* [13, 94, 283, 380] and the approaches discussed in the references therein.

7.5.1.3 MA model

The ARMA model in equation (7.21) reduces to the MA(R) model for the case when $Q = 0$, and is given by:

$$s(t) = \sum_{r=0}^{R} c(r)\, e(t-r) \tag{7.23}$$

In general, it is slightly more difficult to estimate the MA parameters than to estimate the parameters in the AR model, and numerical techniques are often required.

7.5.2 Nonstationary Models

For some time-series, the limitation of assuming a signal is stationary and, therefore, using a stationary model, often results in poor modelling. In such cases the stationary models described above can prove ineffective for some applications. The most common approach to modelling nonstationary processes is to represent the signal in the form of a stationary model, for example those discussed in section §7.5.1, and to represent the model *parameters* as a linear combination of *time-dependent* basis functions. The AR model is a common choice and is the approach applied by Charbonnier *et al*. [63], Liporace [225] and Grenier [130, 131]; Grenier also discusses using basis functions in the ARMA model [130]. Processes possessing nonstationary mean-values are considered by Boudaoud and Chaparro [47] who propose a composite model as the sum of a time-varying function representing the mean, and a zero-mean autoregressive model with time-varying coefficients. The choice of basis functions is dependent on any prior belief of the variation of the parameters [306–308] and, without this knowledge, there exists no general rule for choosing these functions. Proposed classes of basis functions include linear terms [271], polynomials [343], prolate spheroidal sequences [130], wavelets [370], sine and cosines [138], AR processes and Markov processes [378, 379]. Doroslovački and Fan [98] discuss the use of wavelets as basis functions for modelling the impulse response of a linear system given by the superposition integral in equation (3.36), which is a similar approach to using the MA model with time-varying coefficients. Rajan *et al*. [306–308] propose choosing a basis set large enough such that it spans all the information pertaining to the time-varying nature of the AR coefficients, which initially leads to an overparameterised or 'noisy' representation of the coefficients. These overparameterised basis vectors are decomposed, using the Karhunen-Loève Transform [286, 335, 353],[9] into an orthogonal basis set whose principal components are orientated in the direction of maximum energy of the coefficients. The minimal set of basis vectors which give a reasonable representation of the actual time-dependence of the AR coefficients are then identified. Finally, while many of the aforementioned modelling techniques estimate the parameters using the *prediction error criterion*, which seeks to minimise the total squared predic-

[9] Also see Appendix B.1 for further details on the Karhunen-Loève Transform.

tion error, Mrad *et al*. [264] discuss an alternative *polynomial-algebraic* approach based upon skew polynomial operator algebra introduced in section §7.4.3.2, and use it to estimate the parameters of a time-varying ARMA (TVARMA) model whose time-varying coefficients are modelled by a set of basis functions.

Section §7.5.2.2 briefly discusses the complex exponential class of nonstationary parametric signal models, and their relationship to the TVARMA model. Finally, the class of nonstationary models known in the *econometrics* community as *conditional heteroscedastic* (CH) models is outlined in section §7.5.2.3.

7.5.2.1 Time-Varying ARMA model

The general form for a time-varying ARMA (TVARMA) model of a time-series $\{s(t)\}$ is given in terms of a zero-mean *nonstationary* white noise process, $e(t) \sim p_E\left(e(t) \,\middle|\, \sigma^2_{e,t}\right) = \mathcal{N}\left(e(t) \,\middle|\, 0,\ \sigma^2_{e,t}\right)$, as:

$$s(t) = \underbrace{-\sum_{q=1}^{Q_t} b(t,q)\, s(t-q)}_{\text{TVAR terms}} + \underbrace{\sum_{r=0}^{R_t} c(t,r)\, e(t-r)}_{\text{TVMA terms}} \tag{7.24}$$

where $\{b(t,q),\ q \in \mathscr{Q}_t = \{1,\ldots,Q_t\},\ t \in \mathscr{T}\}$ and $\{c(t,r),\ r \in \mathscr{R} = \{1,\ldots,R_t\}; c(t,0) \triangleq 1; t \in \mathscr{T}\}$ are the time-varying model parameters. The special cases of the time-varying AR (TVAR) model and time-varying MA (TVMA) model are analogous in form to the AR and MA models described in equations (7.22) and (7.23) respectively. The time-varying parameters may either be modelled as a linear combination of deterministic time-varying functions, or as an unobserved random process as discussed in the previous sections.

The most general case of the TVARMA model is one where the model parameters are completely uncorrelated at each sample and, therefore, each sample of the signal $s(t)$ would be represented by an ARMA(Q_t, R_t) system or, more explicitly, by $Q_t + R_t$ unknown coefficients. Whilst it may be possible to use such a model, there is little reason to do so, since the parameter space is increased at each sample from 1 to $Q_t + R_t$, causing numerical problems since there isn't enough data, from a single realisation of a process, to allow estimation of the parameters. Therefore, practical cases are ones where the parameters of the TVARMA model are highly correlated such that the parameter space is reduced. Similar model-selection techniques, as discussed in section §7.5.1.2, are applied to time-varying AR process; *e.g.* see Kozin and Nakajima [197].

For reference, Whittle [393] discusses some recursive relationships for linear time-varying predictors of nonstationary processes based on the TVAR process.

7.5.2.2 Unified Approach to Nonstationary Modelling

Complex exponential signal models are a further class of parametric models best suited for transient signals which constitute a specific type of decaying nonstationarity. A complex transient signal $\{x(n),\ n \in \mathcal{N}\}$ may be represented by the complex exponential signal model as [329]:

$$x(n) = \sum_{i=0}^{M-1} A_i\, z_i^n\, e^{j\phi_i} \equiv \sum_{i=0}^{M-1} A_i \left[e^{\alpha_i nT}\right]\, e^{j\omega_i nT}\, e^{j\phi_i}, \qquad \forall\, n \in \mathcal{N} \tag{7.25a}$$

where $z_i = e^{s_i T}$, $s_i = \alpha_i + j\omega_i$, A_i is the carrier amplitude, $\alpha_i < 0$, ω_i is the carrier angular frequency, ϕ_i is the i. i. d. random phase, and T is the sampling interval. The random variable ϕ_i is uniformly distributed over $[0, 2\pi)$. However, when a signal is not 'decaying' over time, the use of such models is questionable [330,331]. Sircar and Syali [331] therefore introduce the complex amplitude modulated (AM) signal model suitable for nonstationary signals encountered in steady-state analysis which may have some repetitive structure, but not any discernible characteristic decay over time, for example, voiced speech signals. The complex AM signal model represents the complex stochastic process $\{x(n),\ n \in \mathcal{N}\}$ by M single-tone AM signals as:

$$x(n) = \sum_{i=0}^{M-1} A_i \left[1 + \mu_i\, e^{j\nu_i nT}\right]\, e^{j\omega_i nT}\, e^{j\phi_i}, \qquad \forall\, n \in \mathcal{N} \tag{7.25b}$$

where A_i, ω_i, ϕ_i and T have the same meaning as in equation (7.25a), μ_i is the modulation index for the sinusoidal modulating signal, and ν_i is the modulating angular frequency. The model parameters are estimated by utilising the accumulated autocorrelation function (AACF)[10] of the modelled signal [329]. Furthermore, the complex frequency modulated (FM) signal model introduced in [330] is shown to be a good parametric approach for unvoiced speech signals and, therefore, a combination of complex AM and FM signal models can be applied for the parametric modelling of continuous speech. The complex random process $\{x(n),\ n \in \mathcal{N}\}$

[10]The accumulated autocorrelation function for a continuous time process $\{x(t),\ t \in \mathcal{T}\}$ is defined as $r_{xx}(\tau) = \int R_{xx}(t,\tau)\, d\tau$, $\forall t \in \mathcal{T}$, where the integral is over $\tau \subset \mathcal{T}$, and a similar definition can be made for discrete time processes [329]. The AACF is also known as the time-averaged correlation function [110].

is represented by the complex FM signal model as:

$$x(n) = \sum_{i=0}^{M-1} A_i \left[e^{j\mu_i \sin(\nu_i nT)} \right] e^{j\omega_i nT} e^{j\phi_i}, \qquad \forall n \in \mathcal{N} \tag{7.25c}$$

where the parameters have the same meaning as in equation (7.25b). The signal model in equation (7.25c) is highly nonlinear and, although estimating the parameters of the model is not an easy task, a procedure is proposed in [330].

In [265] Mukhopadhyay and Sircar demonstrate that the complex exponential, AM, and FM models are actually special cases of the general TVARMA model, where the underlying nonstationary signals can be classified through functional specifications of the movements of the time-varying pole locations. However, as mentioned in section §7.5, a set of basis functions has to be selected for suitable modelling of each type of underlying nonstationary signal, and the optimal choice of basis functions is an issue of continuing research [307, 308].

7.5.2.3 Conditional Heteroscedastic Models

A broad class of *nonstationary* time series modelling used in the *econometric and statistical* communities are conditional heteroscedastic (CH) models, first introduced by Engle [102] in the context of modelling UK inflation. The work by Rajan *et al.* [306] highlights the fact that it is easier to find tractable solutions to parameter estimation when the TVAR coefficients are modelled by a set of basis functions, than when the TVAR variance is modelled as a nonstationary process. In CH models, however, the excitation sequence is modelled as a time-varying *and* state dependent process, while the AR parameters often remain stationary. As an example, the simple autoregressive conditional heteroscedastic (ARCH) has the form:

$$\begin{aligned} y(t) &= -a(0) - \sum_{p=1}^{P} a(p)\, y(t-p) + e(t) \\ e(t) &= h^{1/2}(t)\, \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}\left(\epsilon(t) \,\middle|\, 0,\, 1\right) \end{aligned} \tag{7.26a}$$

where the time-varying *innovation process variance*, $h(t)$, is of the form:

$$h(t) = h\left(e(t-\tau),\, y(t-\tau),\, h(t-\tau),\, 0 < \tau \in \mathbb{Z},\, \mathbf{b}\right) \tag{7.26b}$$

where $\mathbf{b}$ is the variance model parameter vector. The form of this conditional variance depends on the problem, and further details can be found in, for exam-

ple, [44, 45, 102, 275, 323]. These models have the interesting property that the *conditional variance*, defined as the variance of a particular observed data sample given all the past information regarding the state of the system, is stationary, whilst the *unconditional variance* (the usual definition of variance – see section §3.1.2) is nonstationary. The advantage of CH models is that, for a suitable choice of excitation variance, a simple optimisation algorithm can be used to estimate the model parameters, and this therefore facilitates the use of a nonstationary excitation sequence. A related approach by Yokoyama *et al*. [405, 406] is the *inhomogeneous AR model* in which the nonstationary model is a noise-contaminated system of an AR process excited by an input signal modelled by a set of orthogonal basis functions.

7.5.3 Quasi-Stationary Models

As discussed in section §7.5.1, stochastic processes that are *globally* nonstationary, but are approximately *locally* stationary, can be represented by a quasi-stationary model in which the model statistics are stationary over a short segment of time, but globally nonstationary. Quasi-stationary models are, in fact, just a special case of nonstationary models, but are distinguished since they closely resemble stationary modelling techniques. The models used in the deconvolution problem are assumed to be *linear* block stationary, but extensions to *nonlinear* block stationary models is trivial. This model considerably reduces the parameter space in the general TVARMA model by correlating the data over large data segments. The data, $\{s(t)\}$, is partitioned into M contiguous disjoint blocks. The i-th block, $i \in \mathscr{M} = \{1, \ldots, M\}$, begins at sample t_i and has length $T_i = t_{i+1} - t_i$. In this block, $s(t)$ is given by a stationary ARMA model of order $\{Q_i, R_i\}$ and, using the form of equation (7.24), setting $Q_t = Q_i$, $R_t = R_i$, $b(t, q) = b_i(q)$, $q \in \mathscr{Q}_i = \{1, \ldots, Q_i\}$, $c(t, r) = c_i(r)$, $r \in \mathscr{R}_i = \{1, \ldots, R_i\}$, $\sigma_t = \sigma_i$, $\forall t \in \mathscr{T}_i = \{t_i, \ldots, t_{i+1} - 1\} \subset \mathbb{Z}^{T_i}$ and $i \in \mathscr{M}$, yields the block stationary ARMA (BSARMA) process:

$$s(t) = \underbrace{-\sum_{q=1}^{Q_i} b_i(q)\, s(t-q)}_{\text{BSAR terms}} + \underbrace{\sum_{r=0}^{R_i} c_i(r)\, e(t-r)}_{\text{BSMA terms}} \tag{7.27}$$

where $e(t) \sim \mathcal{N}\left(e(t) \,\middle|\, 0,\, \sigma_i^2\right)$, $t \in \mathscr{T}_i$, $i \in \mathscr{M}$. Note that this model implicitly sets the Q_i initial samples of the model in block i to the last Q_i samples in the previous block, $(i-1)$. As for the TVARMA model, the special cases of the block stationary

AR (BSAR) model and block stationary MA (BSMA) model are analogous in form to the usual AR and MA models. The block stationary model is a special case of the basis-function approach discussed in section §7.5.2 in which the time-varying ARMA parameters are piecewise-constant over different blocks or intervals.

7.6 Source Signal Model

This thesis is mainly concerned with the task of speech dereverberation in acoustic environments although, naturally, the blind deconvolution algorithm can be extended to other problems. As such the source model is chosen assuming that the source signal is speech. The analysis of speech signals can be classified into three main groups: those based on (nonstationary) autoregressive models [18, 117, 120, 130, 137, 138, 221, 237, 269, 271, 378, 379, 381], those based on Fourier analysis or sinusoidal decomposition [10, 58, 113, 125, 244, 248, 301], and AM-FM or modulated signal models [309, 330, 331]. The sensitivity of the model to the nature of the underlying signal structure should decrease if additional information inherent in the nonstationarity of the process is utilised; as such, in this work, speech signals are modelled using autoregressive rather than sinusoidal or Fourier decompositions. In section §7.5.3, the BSARMA model is introduced, and it is noted in section §7.5.1.2 that the AR model was particular useful for modelling speech signals over short periods of time. Moreover, the more general BSARMA model leads to a complicated expression for parameter estimation which does not lend itself to closed form analytical solutions, although numerical Bayesian methods can be deftly applied [118]. As such, since an investigation of nonstationarity requires analytical solutions, the source signal, $s(t)$, assumed to be speech, will be modelled by the BSAR model:[11]

$$s(t) = -\sum_{q=1}^{Q_i} b_i(q)\, s(t-q) + e(t) \tag{7.28}$$

where $e(t) \sim \mathcal{N}\left(e(t) \,\middle|\, 0,\, \sigma_i^2\right)$, $\sigma \in \mathbb{R}^+$, $t \in \mathscr{T}_i$, $i \in \mathscr{M}$. A variation on this model is introduced in Chapter 9 in which the source signal is modelled over a particular frequency range by a spectral AR model in the form of equation (7.28).

[11] BSAR processes are also known as *piecewise constant autoregressive processes* [283, 298].

7.6.1 Synthetic Source Signals

Throughout this dissertation, the properties of the proposed models will be demonstrated by using source signals which are either real data, or synthetically generated data. The algorithms developed must necessarily provide useful estimates for real data, whilst synthetic data is useful to test the algorithm in 'ideal conditions' where the data is *known* to fit the BSAR model exactly. The following two classes of synthetic data are used:

BSAR(2) **Linear Variation:** The signal is generated as a second-order block stationary AR process, with the phase and magnitude of the pole locations changing linearly with data blocks.

BSAR(12) **Random Variation** The signal is generated as a 12th-order block stationary AR process, with the parameters in each block equated to the least-squares parameter estimates for an AR model of segments of real data.

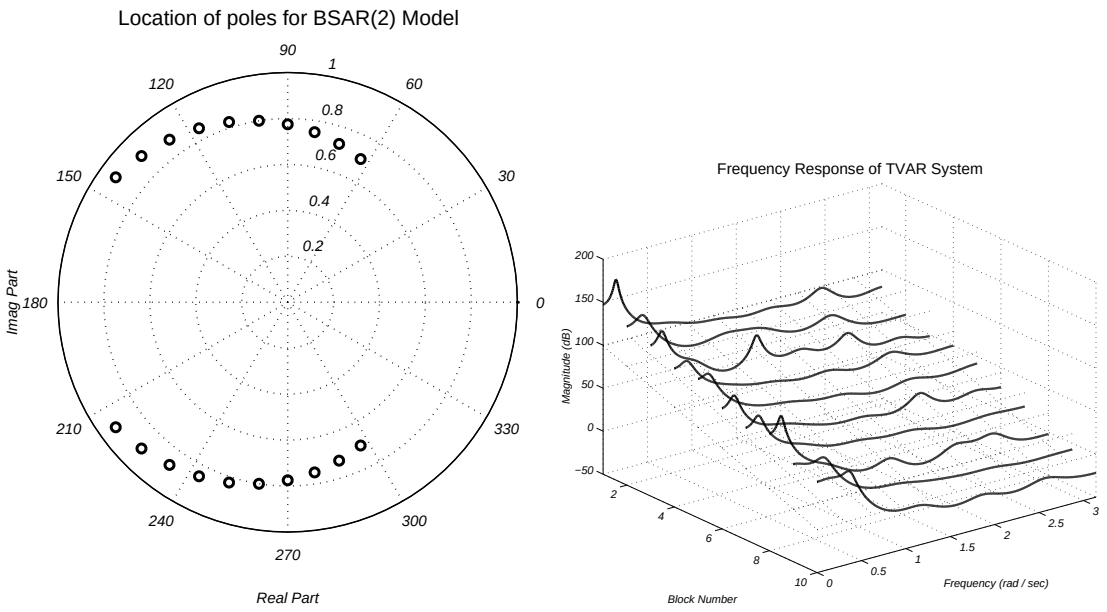
These models are discussed in further detail in the next two sections.

7.6.1.1 BSAR(2) Linear Variation Synthetic Signal Model

The first synthetic source signal is the block stationary AR(2) model, which is useful since it allows the effect of the source signal on the estimation of the parameters for a second-order channel to be observed in full; higher order models are difficult to visualise, and useful results cannot easily be plotted. Given this source model, within each block the system relates its input, $e(t)$, to its output, $s(t)$, by $e(t) = s(t) - 2r_i \cos\theta_i\, s(t-1) + r_i^2\, s(t-2)$, where r_i and θ_i are, respectively, the magnitude and phase of the single complex pole-pair at $z_i = r_i\, e^{\pm j\theta_i}$, $i \in \mathscr{M}$. Speech is assumed to be varying at a rate in which the frozen-state approximation of system poles is appropriate (see section §7.4.3.1). The location of this pole changes with block i, and is characterised by the following variations in radius and phase:

$$\left.\begin{aligned} r_i &= r_{min} + \frac{r_{max} - r_{min}}{M} i \\ \theta_i &= \theta_{min} + \frac{\theta_{max} - \theta_{min}}{M} i \end{aligned}\right\} \quad i \in \{1, \ldots, M\} \tag{7.14}$$

where M is the number of data blocks, and $\{r_{min},\, \theta_{min}\}$, $\{r_{max},\, \theta_{max}\}$ determine the initial and final locations of the time-varying pole. Using the following val-



(a) The trajectory of the time-varying pole for the BSAR(2) linear variation synthetic signal model.

(b) Frequency response within each data block for the BSAR(12) random variation synthetic signal model.

Figure 7.7: *Frequency characteristics for the synthetic source models.*

ues for the parameters, $M = 10$, $(r_{min}, r_{max}) = (0.70, 0.95)$, and $(\theta_{min}, \theta_{max}) = (0.35\pi, 0.85\pi)$, the trajectory of the pole is shown in Figure 7.7(a).

7.6.1.2 BSAR(12) Random Variation Synthetic Signal Model

The second synthetic source signal reflects the statistical properties possessed by a speech signal and, in this case, the BSAR process is 12th-order. The parameters in each block are equated to the least-squares estimate of the parameters for an AR model of a block of real speech sampled at 11.025 kHz. This speech signal is taken from the first 12 seconds of *Suzanne Vega's 'Tom's Diner'*. The parameters in block i are estimated by modelling the actual data in discontiguous blocks, ensuring the resulting parameter estimates are nonstationary, by a AR(Q) model using the autocorrelation method [237, 353]. This ensures that the parameters lie within the stability region and, therefore, that the synthesised signal is stable. The synthetic data is then generated, in contiguous blocks, using these parameter values, and a typical resulting frequency response is shown in Figure 7.7(b).

7.7 Bayesian Parameter Estimation

Deconvolution of a system requires either an estimate of the original signal, or an estimate of the channel itself and, in either case, some unknown signal or system must be estimated. However, whilst, in the past, it was conventional to consider an unknown signal as a *random* process, it is instinctive, perhaps more often than ought, to consider the channel as an unknown *deterministic* process whose characteristics can be estimated using traditional *point-estimates* as, for example, in the *method of least-squares*. The channel should, in fact, be considered as an unknown stochastic system and, for parametric models, this is easily achieved by considering the system parameters and excitation sequence as random variables. This stochastic view of the system implies that point-estimates of the channel parameters are no longer satisfactory and, although it may be the case that a point-estimate is required to design an effective deconvolution procedure, the method of obtaining such an estimate must accommodate the idea that the underlying system is stochastic.

Encapsulating the stochastic nature of both the system itself and the input–output signals is deftly achieved using the Bayesian methodology, in which the entire system is considered as a stochastic process. However, the closed system cannot be considered in general terms since, in such a case, the Bayesian method becomes intractable. To apply the Bayesian method, constraints must be incorporated on the system in terms of *prior knowledge* regarding the form of model of the system. The problem of selecting a model for a system is a subject in itself, but it is important to note that when a model is chosen, some assumption regarding the system has been made. All too often, the Bayesian methodology can appear to yield results with only minor intervention on the part of the Engineer, and this perception is often influenced by the lack of clearly stated assumptions. The Bayesian approach, however, elegantly records and incorporates assumptions regarding the problem. If there are properties of a system which are believed to be known, an 'impulsive' prior can be assigned to those properties, which means there is a zero probability of the system having a property other than that which is believed it should have.

This section discusses the Bayesian approach to parameter estimation, the form in which the *prior information* is expressed, and the derivation of some results that are used throughout the rest of this dissertation.

7.7.1 Bayes' Theorem

If a particular event, $\mathscr{A}$, has occurred, then the probability of a second event, $\mathscr{B}$, can be calculated given knowledge of $\mathscr{A}$, and is denoted by $\Pr(\mathscr{B} \mid \mathscr{A})$. This probability can be calculated even if event $\mathscr{A}$ has not been observed, or is unobservable: it must merely be an event which can be defined within the context of the system. Moreover, if an event, $\mathscr{B}$, *has* occurred, then the probability of that event, assuming that an event, $\mathscr{A}$, previously occurred, is also given by $\Pr(\mathscr{B} \mid \mathscr{A})$ and can be calculated retrospectively. This gives rise to the *likelihood* function which is the probability of realising a particular state of a system, $\mathbf{x}$, *assuming* the system parameters takes on the value $\boldsymbol{\theta}$, and an underlying model, $\mathcal{I}$, is known. The maximum-likelihood (ML) approach to parameter estimation then seeks to maximise $\Pr(\mathbf{x} \mid \boldsymbol{\theta}, \mathcal{I})$ w. r. t. $\boldsymbol{\theta}$, under the pretence that it is maximising the probability of the system state, $\mathbf{x}$, occurring given a particular set of system parameters, $\boldsymbol{\theta}$, and a particular system model, $\mathcal{I}$. The set of parameters which yield the maximum probability of the current state can then be taken as a *point-estimate* of the system parameters, known as the maximum-likelihood estimate (MLE). Under certain conditions, it can be shown that the MLE is equivalent to the *minimum mean-square estimate* (MMSE) [283, 335, 353].

However, it is equally intuitive, if not more so, to maximise the probability of a particular set of system parameters, $\boldsymbol{\theta}$, given the state of the system, $\mathbf{x}$. This is the *posterior probability*, $\Pr(\boldsymbol{\theta} \mid \mathbf{x}, \mathcal{I})$, since is summarises the knowledge of the system parameters *after* the state of the system has been observed. The posterior probability and the likelihood are related through the prior knowledge of the system, $\Pr(\boldsymbol{\theta} \mid \mathcal{I})$, by Bayes' theorem:

$$\Pr(\boldsymbol{\theta} \mid \mathbf{x}, \mathcal{I}) = \frac{\Pr(\mathbf{x} \mid \boldsymbol{\theta}, \mathcal{I}) \ \Pr(\boldsymbol{\theta} \mid \mathcal{I})}{\Pr(\mathbf{x} \mid \mathcal{I})} \tag{7.29a}$$

This can be generalised to deal with continuous random variables by replacing probabilities by probability density functions[12] to give the modified version of

[12] The probability of a random variable X taking on a value between x_1 and x_2, given that a second random variable Y takes on the value y, is given by:

$$\Pr(x_1 < X < x_2 \mid Y = y) = \int_{x_1}^{x_2} p_{X \mid Y}(x \mid y) \ \mathrm{d}x$$

where $p_{X \mid Y}(x \mid y)$ is the probability density function (pdf) of x given y.

Bayes' theorem:

$$\mathrm{p}\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x}, \mathcal{I}\right) = \frac{\mathrm{p}\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta}, \mathcal{I}\right)\, \mathrm{p}\left(\boldsymbol{\theta} \,\middle|\, \mathcal{I}\right)}{\mathrm{p}\left(\mathbf{x} \,\middle|\, \mathcal{I}\right)} \tag{7.29b}$$

The term $\mathrm{p}\left(\mathbf{x} \,\middle|\, \mathcal{I}\right)$ is the *evidence* and, although it is usually regarded as a normalising constant, it is of interest for model selection (see section §8.3).

7.7.2 Prior Distributions

Two components in the systems discussed throughout this dissertation are:

1. stationary zero mean white Gaussian noise (WGN) sources, and
2. infinite impulse response (IIR) linear time-invariant (LTI) finite-order filters.

Together these define the autoregressive process defined in equation (7.28) and IIR filters (7.5) which, respectively, model the source signals and the channel responses. The prior distributions in the Bayesian model represent the prior belief of these components before the state of the system (the observed data) is known. If there is complete ignorance of which values the parameter are likely to take, then priors which do *not* influence the posterior distribution should be used, and are known as *non-informative* priors. In such a case, the posterior distribution essentially reduces to the likelihood function. In order to obtain analytical results, the prior distribution must usually be expressed as a smooth continuous function across the entire parameter space and, as such, the class of improper *Jeffrey's priors*[13] are often used. However, improper priors do not always facilitate model comparison or the drawing of samples from a distribution. A more flexible approach is to choose proper prior distributions which are close to being uninformative, and which depend on *hyperparameters* that must somehow be chosen. In some cases, these hyperparameters can themselves be considered as random variables, and *hierarchical modelling* can be applied so that eventually the choice of some specified hyper-hyperparameters do not greatly influence the posterior distribution. It should be noted, however, that uninformative priors which are representative of a system can sometimes be hard to find. A useful class of proper priors, in which a suitable choice of hyperparameters can render the prior essentially uninformative, is known as the class of *conjugate priors*. These lead to the posterior distribution taking the

[13] Jeffrey's priors are improper in the sense that they cannot be normalised: $\int_{\Theta} \mathrm{p}(\theta)\, \mathrm{d}\theta \to \infty$.

same form as the prior, and allow more analytically-tractable distributions to be obtained.[14] The following subsections discuss the details of such prior distributions which prove useful in the analysis of the proposed models.

7.7.2.1 Prior distribution on AR coefficients

In the case when a process is modelled by a real stable minimum-phase AR process of order P with parameters $\mathbf{a}$ and excitation variance σ^2, the parameter vector, $\mathbf{a}$, should ideally only take on values which lie in the *stability domain*. However, the terms in the likelihood function for AR parameters are usually in the form of a Gaussian distribution (see [49, 238]) and, in order to obtain analytically tractable results, a Gaussian prior is placed on the parameter values: $\mathbf{a} \,|\, \sigma^2 \sim \mathcal{N}\left(\mathbf{a} \,\middle|\, \mathbf{0}_P,\ \sigma^2\,\delta^2\,\mathrm{I}_P\right),\ \delta \in \mathbb{R}^+$. The hyperparameter, δ, expresses the variance of this conditional prior distribution, and the prior on $\mathbf{a}$ becomes uninformative as $\delta \to \infty$. $\mathrm{I}_P \in \mathbb{R}^{P\times P}$ is the identity matrix.

7.7.2.2 Prior distribution for the Excitation Variance

A standard prior for application to scale parameters, such as variances, is the inverse-Gamma density with the form [35]:

$$p\left(\sigma^2 \,\middle|\, \alpha, \beta\right) = \mathcal{IG}\left(\sigma^2 \,\middle|\, \alpha,\ \beta\right) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\beta}{\sigma^2}\right] \mathbb{I}_{(0,+\infty]}\left(\sigma^2\right) \qquad (7.30)$$

The non-informative improper Jeffrey's prior is a special case of the inverse-Gamma density: $p\left(\sigma^2\right) = \sigma^{-2}$.

7.7.3 Posterior Distribution for Source Signal

In this section, the posterior distribution for a general system driven by the source signal modelled as a BSAR process, as discussed in section §7.6, is derived. Consider the general system in Figure 7.8, where the output of the system, $x(t)$, is related to the input, $s(t)$, by $x(t) = f(s(t), \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the parameters of the system model denoted by $\mathcal{I}$. The source signal, $s(t)$, is given by equation (7.28), where the excitation $e(t) \sim \mathcal{N}\left(e(t) \,\middle|\, 0,\ \sigma_i^2\right)$, $\sigma_i \in \mathbb{R}^+$, $t \in \mathscr{T}_i$, $i \in \mathscr{M}$. Within that block, $s(t)$ is assumed to be given by a stationary AR model of order Q_i. Therefore, the

[14] Note, however, that the form of these conjugate priors is constrained by the form of the likelihood function [35].

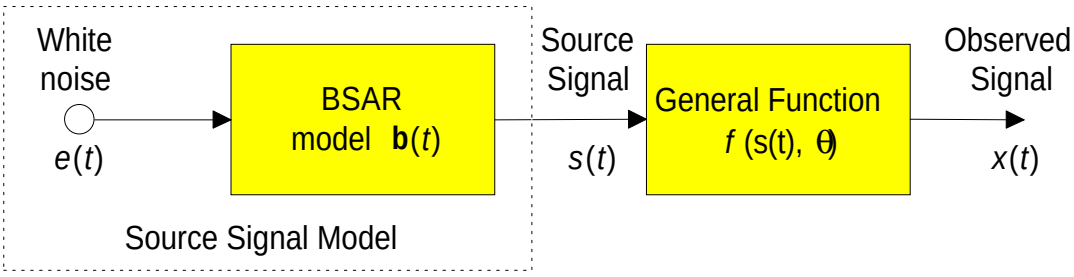


Figure 7.8: *A general system* $x(t) = f(s(t), \boldsymbol{\theta})$ *driven by a BSAR process.*

excitation samples in block $i \in \mathscr{M}$ may be written as $\mathbf{e}_i = \mathbf{s}_i + \mathbf{S}_i \mathbf{b}_i$, where $\mathbf{e}_i$ is a vector of samples $[\mathbf{e}_i]_{t-t_i+1} = \mathbf{e}(t)$, $t \in \mathscr{T}_i$ and, similarly, $[\mathbf{s}_i]_{t-t_i+1} = \mathbf{s}(t)$, $t \in \mathscr{T}_i$, $\mathbf{b}_i$ is a vector of parameters $[\mathbf{b}_i]_q = b_i(q)$, $q \in \mathscr{Q}_i = \{1, \ldots, Q_i\}$, and the data matrix $[\mathbf{S}_i]_{t-t_i+1,q} = \mathbf{s}(t-q)$, $t \in \mathscr{T}_i$, $q \in \mathscr{Q}_i$. The probability distribution for the excitation in block i is therefore given by:

$$\mathbf{e}_i \sim p_{\mathbf{e}_i}\left(\mathbf{e}_i \,\middle|\, \sigma_i^2\right) = \mathcal{N}\left(\mathbf{e}_i \,\middle|\, \mathbf{0}_{T_i},\, \sigma_i^2 \mathbf{I}_{T_i}\right) = \frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{-\frac{\mathbf{e}_i^T \mathbf{e}_i}{2\sigma_i^2}\right\} \tag{7.31}$$

The probability *chain rule* [286] identity is given by:

$$p\left(\mathbf{s} \,\middle|\, \mathcal{I}\right) = p\left(\mathbf{s}_i,\, i \in \mathscr{M} \,\middle|\, \mathcal{I}\right) = p\left(\mathbf{s}_1 \,\middle|\, \mathcal{I}\right) \prod_{i=2}^{M} p\left(\mathbf{s}_i \,\middle|\, \mathbf{s}_{i-1}, \ldots, \mathbf{s}_1, \mathcal{I}\right) \tag{7.32a}$$

which, since the block stationary AR process depends only on the previous Q_i outputs, reduces to:

$$p\left(\mathbf{s} \,\middle|\, \mathcal{I}\right) = p\left(\mathbf{s}_1 \,\middle|\, \mathcal{I}\right) \prod_{i=2}^{M} p\left(\mathbf{s}_i \,\middle|\, \mathbf{s}_{i-1}, \mathcal{I}\right) \tag{7.32b}$$

The Jacobian[15] $\mathscr{J}(\mathbf{s}_i, \mathbf{e}_i)$ is unity since the transformation is linear. Denoting $\boldsymbol{\sigma} = \{\sigma_i^2,\, i \in \mathscr{M}\}$, $\mathbf{b} = \{\mathbf{b}_i,\, i \in \mathscr{M}\}$, and $\boldsymbol{\psi} = \{\boldsymbol{\sigma}, \mathbf{b}\}$, it follows that the likelihood function for the source signal, $\mathbf{s}_i$, in block $i \in \mathscr{M}_{\{-1\}} \triangleq \mathscr{M} - \{1\}$ is given by:

$$p\left(\mathbf{s}_i \,\middle|\, \mathbf{s}_{i-1}, \boldsymbol{\psi}, \mathcal{I}\right) = \frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{-\frac{(\mathbf{s}_i + \mathbf{S}_i\mathbf{b}_i)^T(\mathbf{s}_i + \mathbf{S}_i\mathbf{b}_i)}{2\sigma_i^2}\right\} \tag{7.33}$$

[15]The Jacobian for the transformation $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is $\mathscr{J}(\mathbf{y}, \mathbf{x}) = \left|\frac{\partial \mathbf{f}^T}{\partial \mathbf{x}}\right|$

where $i \in \mathscr{M}_{\{-1\}}$, and $\mathscr{M}_{\{-1\}}$ denotes the set $\mathscr{M}$ not including the element 1. The distribution $p\left(s_1 \,\middle|\, \mathcal{I}\right)$ can be decomposed into the form:

$$p\left(s_1 \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right) = p\left(\{\hat{s}_1, s_0\} \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right) = p\left(\hat{s}_1 \,\middle|\, s_0, \boldsymbol{\psi}, \mathcal{I}\right)\, p\left(s_0 \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right)$$

where s_0 are the initial values of the signal s. The first term $p\left(\hat{s}_1 \,\middle|\, s_0, \boldsymbol{\psi}, \mathcal{I}\right)$ has an identical form to equation (7.33) whilst, since the data is assumed zero mean and Gaussian, the second term is given by $p\left(s_0 \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right) = \mathcal{N}\left(s_0 \,\middle|\, \mathbf{0},\, \sigma_1^2 M_{s_0}\right)$, where M_{s_0} is the covariance matrix for P samples of data with unit variance [49]. However, since M_{s_0} is a function of the parameters $\mathbf{b}_1$, it can be seen that the exact likelihood function for $p\left(s_1 \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right)$ becomes highly nonlinear and the posterior distribution becomes intractable to analytic solution. Nevertheless, if $T_1 \gg Q_1$, which is often the case with audio signals, it is common practice to make the approximation:

$$p\left(s_1 \,\middle|\, \boldsymbol{\psi}, \mathcal{I}\right) \approx p\left(\hat{s}_1 \,\middle|\, s_0, \boldsymbol{\psi}, \mathcal{I}\right)$$

The simplification is implicitly assumed throughout the rest of the derivation of the posterior distribution and, for clarity, $\hat{s}_1$ and s_1 are considered synonymous. The Jacobian $\mathscr{J}(\mathbf{x}, s)$ is given by:

$$\mathscr{J}(\mathbf{x}, s) = \operatorname{abs}\left|\frac{\partial f(s, \boldsymbol{\theta})}{\partial s}\right| \tag{7.34}$$

and, therefore, using (7.32b) the likelihood function for the observed signal is:

$$p\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\phi}, \mathcal{I}\right) = \frac{1}{\mathscr{J}(\mathbf{x}, s)} \prod_{i=1}^{M} \frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{-\frac{(s_i + S_i \mathbf{b}_i)^T (s_i + S_i \mathbf{b}_i)}{2\sigma_i^2}\right\} \tag{7.35}$$

where it is noted $s(t) \equiv s(t, \mathbf{a}, \mathbf{x})$ is given by the relationship $\mathbf{x} = f(s, \boldsymbol{\theta})$, and where the parameter vector $\boldsymbol{\phi} \triangleq \{\boldsymbol{\tau}, \boldsymbol{\Xi}, \boldsymbol{\delta}, \boldsymbol{\nu}, \boldsymbol{\gamma}\}$ contains the vector of *changepoints* (or boundaries) of the data blocks, $\boldsymbol{\tau} = \{t_i,\, i \in \mathscr{M}\}$, the vector of model orders, $\boldsymbol{\Xi} = \{Q_i,\, i \in \mathscr{M}\}$, and the vectors of hyperparameters, $\boldsymbol{\delta} = \{\delta_i,\, i \in \mathscr{M}\}$, $\boldsymbol{\nu} = \{\nu_i,\, i \in \mathscr{M}\}$, and $\boldsymbol{\gamma} = \{\gamma_i,\, i \in \mathscr{M}\}$ as defined in the assigned priors below. Applying Bayes' rule, the posterior pdf for the unknown parameters $\{\boldsymbol{\theta}, \boldsymbol{\psi}\}$ becomes:

$$p\left(\boldsymbol{\theta}, \boldsymbol{\psi} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) = \frac{p\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\phi}, \mathcal{I}\right)\, p\left(\boldsymbol{\theta}, \boldsymbol{\psi} \,\middle|\, \boldsymbol{\phi}, \mathcal{I}\right)}{p\left(\mathbf{x} \,\middle|\, \boldsymbol{\phi}, \mathcal{I}\right)} \tag{7.36a}$$

where, for the moment, the parameter vector $\boldsymbol{\phi}$ is assumed to be known. As discussed in section §7.7.2, suitable choices for the priors on the AR coefficients and the excitation variances for the source model are Gaussian and inverse-Gamma distributions respectively. Furthermore, assuming that the time-varying model parameters, $\{\mathbf{b}_j,\ \sigma_j\}$, are independent between data blocks, then the assigned priors are $\mathbf{b}_j \,|\, \sigma_j^2 \sim \mathcal{N}\left(\mathbf{b}_j \,\middle|\, \mathbf{0}_{Q_j},\ \sigma_j^2\, \delta_j^2\, \mathrm{I}_{Q_j}\right),\ \delta_j > 0$, and $\sigma_j^2 \sim \mathcal{IG}\left(\sigma_j^2 \,\middle|\, \frac{\nu_j}{2},\ \frac{\gamma_j}{2}\right)$, $j \in \mathscr{M}$. Ergo, equation (7.36a) reduces to:

$$p\left(\boldsymbol{\theta},\ \boldsymbol{\psi} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right) = \frac{p\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta},\ \boldsymbol{\psi},\ \boldsymbol{\phi},\ \mathcal{I}\right)\ p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi},\ \mathcal{I}\right)\ p\left(\mathbf{b} \,\middle|\, \boldsymbol{\sigma},\ \boldsymbol{\phi},\ \mathcal{I}\right)\ p\left(\boldsymbol{\sigma} \,\middle|\, \boldsymbol{\phi},\ \mathcal{I}\right)}{p\left(\mathbf{x} \,\middle|\, \boldsymbol{\phi},\ \mathcal{I}\right)} \tag{7.36b}$$

Since it is only of interest to estimate the channel parameters, $\boldsymbol{\theta}$, the source signal parameters are a *nuisance*, and it would be useful to obtain a particular estimate of $\boldsymbol{\theta}$ which is not conditional on the BSAR parameters or excitation variances. These *nuisance* parameters can be *marginalised* using the probability relationship:

$$p_x\left(\mathbf{x}\right) = \int_{\mathbf{Y}} p_{x,y}\left(\mathbf{x},\ \mathbf{y}\right)\ \mathrm{d}\mathbf{y} \tag{7.37}$$

In this case, marginalising equation (7.35) w. r. t. the nuisance parameters $\{\mathbf{b}_j,\ \sigma_j\},\ \forall j \in \mathscr{M}$, with the integrals,

$$p\left(\boldsymbol{\theta},\ \boldsymbol{\sigma} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right) = \int_{\mathbb{R}^{Q_1}} \cdots \int_{\mathbb{R}^{Q_M}} p\left(\boldsymbol{\theta},\ \boldsymbol{\psi} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right)\ \mathrm{d}\mathbf{b}_M \ldots \mathrm{d}\mathbf{b}_1 \tag{7.38a}$$

and

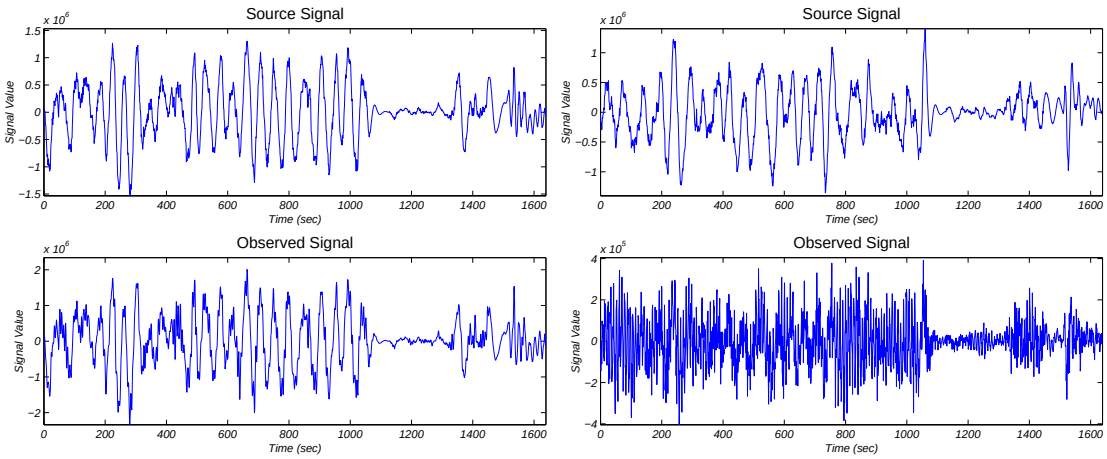
$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right) = \int_0^\infty \cdots \int_0^\infty p\left(\boldsymbol{\theta},\ \boldsymbol{\sigma} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right)\ \mathrm{d}\sigma_M^2 \ldots \mathrm{d}\sigma_1^2 \tag{7.38b}$$

yields, as shown in Appendix D, the posterior density:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right) \propto \frac{p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi},\ \mathcal{I}\right)}{\mathscr{J}\left(\mathbf{x}, \mathbf{s}\right)} \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T \mathrm{S}_i \left(\mathrm{S}_i^T \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right)^{-1} \mathrm{S}_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|\mathrm{S}_i^T\, \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right|^{\frac{1}{2}}} \tag{7.39a}$$

where $R_i = \frac{T_i + \nu_i + 1}{2}$. If there is only one data block, this expression reduces to:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x},\ \boldsymbol{\phi},\ \mathcal{I}\right) \propto \frac{p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi},\ \mathcal{I}\right)}{\mathscr{J}\left(\mathbf{x}, \mathbf{s}\right)} \frac{\left\{\gamma + \mathbf{s}^T \mathbf{s} - \mathbf{s}^T \mathrm{S} \left(\mathrm{S}^T \mathrm{S} + \delta^{-2} \mathrm{I}_Q\right)^{-1} \mathrm{S}^T \mathbf{s}\right\}^{-R}}{\left|\mathrm{S}^T\, \mathrm{S} + \delta^{-2} \mathrm{I}_Q\right|^{\frac{1}{2}}} \tag{7.39b}$$



(a) Input and output signals for the second-order LTI IIR filter shown in Figure 7.10(a).

(b) Input and output signals for the 8th-order LTI IIR filter shown in Figure 7.10(b).

Figure 7.9: *Source and distorted signals after the filtering shown in Figure 7.10.*

where $R = \frac{\nu+T+1}{2}$, the subscript i has been dropped for clarity, since there is only a single data block, and assuming a suitable definition of the vector $\mathbf{s}$ and data matrix $\mathbf{S}$. These two expressions are used throughout this thesis.

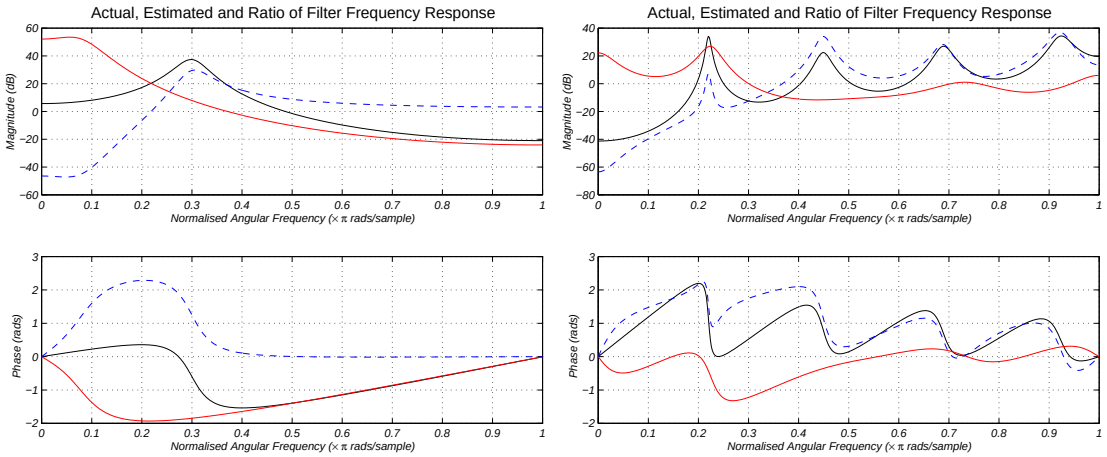
7.8 Performance of Stationary Analysis

To provide motivation for incorporating extra degrees of freedom associated with a nonstationary system, consider the results obtained if a nonstationary system is modelled as stationary one. The two examples introduced below are investigated extensively in the forthcoming chapters. The source signal, $s(t)$, is the synthetic block stationary process discussed in section §7.6.1.2, and the observed signal is a LTI filtered version of $s(t)$. The source and observed signals, for two different LTI filters, $\mathscr{A}_1$ and $\mathscr{A}_2$, are shown in Figure 7.9. The actual filter responses are shown as *black* lines in Figure 7.10. Now, if the entire system is assumed to be stationary, where the system is modelled using the expression in equation (7.39b), and the source model parameters have been marginalised, then the estimate of the filter parameters given the observed data, and assuming the correct model orders are known, yields the estimated spectra shown as *red* lines in Figure 7.10. The

ratio of the actual filter's spectrum and the estimated filter's spectrum are shown in the *dotted blue* lines. Furthermore, the actual and estimated pole locations for the filter are shown in Figure 7.11. It can be seen that the estimate of the filter when the system is modelled as a stationary process is so poor that the spectrum of the equalised filter response is almost as bad as the original filter's spectrum and, therefore, the equalised speech will be as distorted as the unequalised speech. The techniques discussed in this thesis will improve on the estimate considerably.

7.9 Chapter Summary

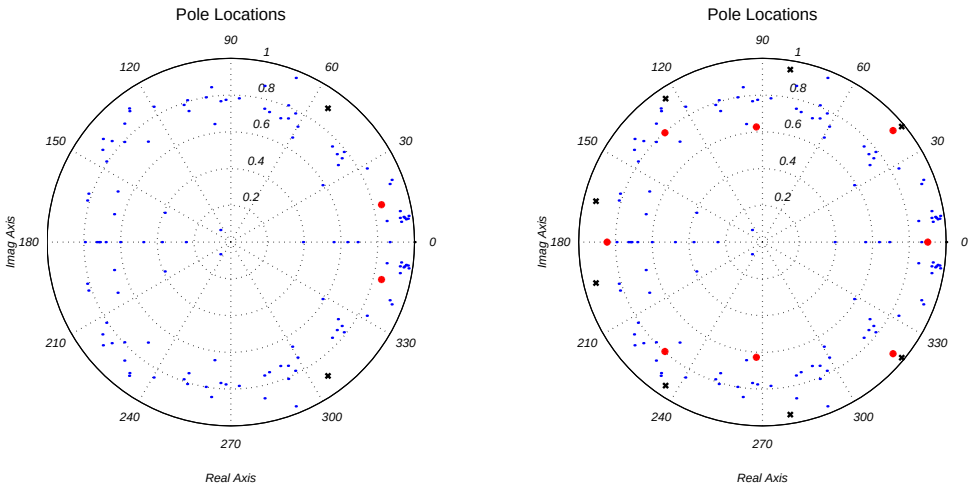
Single channel blind deconvolution is formally introduced in this chapter. Linear input–output modelling that permit systems and nonstationary stochastic processes to be modelled in terms of finite-order linear models, or time-series, is reviewed. Some of these models are used in the next couple of chapters to develop a solution for the problem. There exists a plethora of nonstationary linear and nonlinear input–output and time-series models, each of which is appropriate for different nonstationary systems. The TVARMA model is very general and, as discussed, is a superset of models such as the complex exponential, AM, and FM models. However, practical application of the TVARMA model usually requires a constraint on its parameter space, often in terms of some prior knowledge regarding a particular system. The purpose of this dissertation is not to thoroughly investigate the properties of all these models – this has been done competently elsewhere (see, for example, most of the references in this chapter). Instead, this research is concentrating on how *nonstationarity* can provide additional degrees of freedom which allow strong requirements on prior information to be relaxed. As such, models that have not been investigated further in this research are not *necessarily* unsuitable for blind deconvolution: the models investigated here have been so, since their properties allow the effect of nonstationarity to be more easily observed. The Bayesian paradigm is introduced as a means of parameter estimation, and a theoretical development has been presented which will be used in the forthcoming chapters.



(a) Second-order LTI IIR filter.

(b) 8th-order LTI IIR filter.

Figure 7.10: *The frequency responses of typical LTI IIR filter are shown in black. In each case, the MLE estimate of the filter given the observed data when the system is assumed stationary is shown in red, and the ratio of this estimate to the true response is shown in blue.*



(a) Second-order LTI IIR filter.

(b) 8th-order LTI IIR filter.

Figure 7.11: *The pole locations of the actual and estimated filters for the filters shown in Figure 7.10. The true positions are denoted by a black cross, the estimated positions by a red dot, and the pole locations of the source signal for all the data blocks are shown in blue for comparison.*

8

Bayesian Blind Deconvolution

IN section §1.4, it was postulated that utilisation of the nonstationary properties of a particular system helps toward finding solutions which would otherwise yield unsatisfactory results if the system was modelled as a stationary system. Moreover, it was postulated that if the components of a system have statistics which vary at different rates, then identification of each component should be possible. This basic idea is used by Law and Nguyen [209] for de-blurring images in which multiple frames of a recorded image are used, based on the assumption that the blurring or point spread function is time-varying. In this chapter, a composite model, presented in [161] and [163], is proposed which confirms these propositions, and the properties of this model are investigated. This investigation provides the framework for the generalised model discussed in Chapter 9.

As outlined in section §7.6, the source signal, $s(t)$, is modelled by a BSAR process given by (7.28). The distortion operator, $\mathscr{A}$, is modelled by a LTI IIR filter of order P, such that the observed signal, $\{x(t)\}$, is given in terms of $\{s(t)\}$ by:

$$x(t) = -\sum_{p=1}^{P} a(p)\, x(t-p) + s(t), \quad t \in \mathbb{Z} \tag{8.1}$$

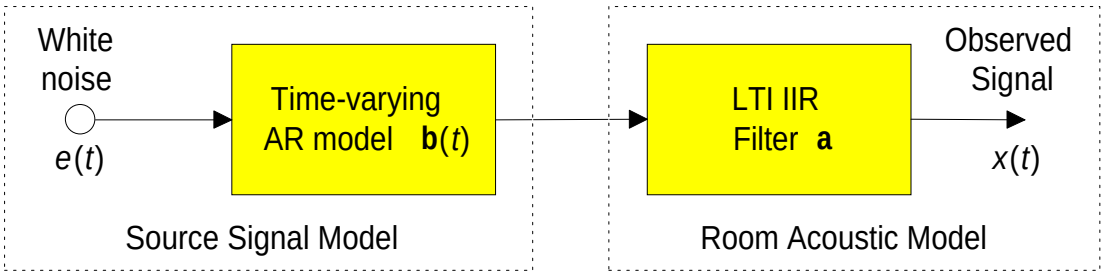


Figure 8.1: *Simplified signal model of the blind deconvolution problem.*

where $\mathbf{a} = \{a(p)\, p \in \mathscr{P} = \{1, \ldots, P\}$ are the model parameters, and P is the number of poles. This system model is shown in Figure 8.1. A Bayesian solution for parameter estimation of the distortion operator, $\mathscr{A}$, is presented in section §8.2. First, however, it is insightful to consider a simple intuitive histogram technique of estimating the channel parameters.

8.1 Exploration of Parameter Space for Unconstrained Channel Model

The parameters for the distortion filter, $\mathscr{A}$, can be estimated using a simple histogram technique introduced in [333,334], developed further in [160], and recently implemented in real time applications [352]. Although this method lacks robustness, it gives considerable insight into the underlying physical process that occurs in the proposed blind deconvolution technique. Consider, in data block i, the output, $x(t)$, of the BSAR-AR model, shown in Figure 8.1, as a windowed version of an infinite stationary data sequence, $x_i(t)$, given by:

$$\left.\begin{aligned} x_i(t) &= -\sum_{p=1}^{P} a(p)\, x_i(t-p) + s_i(t) \\ s_i(t) &= -\sum_{q=1}^{Q_i} b_i(q)\, s_i(t-q) + e_i(t) \end{aligned}\right\} \quad t \in \mathscr{T} \subset \mathbb{Z} \tag{8.2}$$

where $e_i(t) \sim \mathcal{N}\left(e_i(t) \,\middle|\, 0,\, \sigma_i^2\right)$, such that $x(t) = x_i(t), \forall t \in \mathscr{T}_i \subset \mathscr{T}$. The power spectrum, $\mathcal{P}_i(\omega)$, of the signal $x_i(t)$ can be obtained by taking the square modulus

of the *discrete Fourier transform* (DFT) to give:

$$\begin{aligned}\mathcal{P}_i(\omega) &= \frac{1}{\left|1+\sum_{p=1}^{P} a(p)\, e^{-j\omega p}\right|^2}\,\frac{\sigma_i^2}{\left|1+\sum_{q=1}^{Q_i} b_i(q)\, e^{-j\omega q}\right|^2} \\ &\equiv \frac{1}{\prod_{p=1}^{P}\left|1-r_{\mathbf{a}}(p)\, e^{-j\omega}\right|^2}\,\frac{\sigma_i^2}{\prod_{q=1}^{Q_i}\left|1-r_{\mathbf{b}_i}(q)\, e^{-j\omega}\right|^2}\end{aligned} \tag{8.3a}$$

which can be written in the form:

$$\mathcal{P}_i(\omega) = \frac{\sigma_i^2}{\left|1+\sum_{k=1}^{P+Q_i} c_i(k)\, e^{-j\omega k}\right|^2} \equiv \frac{\sigma_i^2}{\prod_{k=1}^{P+Q_i}\left|1-r_i(k)\, e^{-j\omega}\right|^2} \tag{8.3b}$$

where $c_i(k) = [a(p) \star b_i(q)]_k$ are the coefficients of the combined AR$(P+Q_i)$ process. The roots $\mathbf{r}_i \equiv [\,\mathbf{r}_i(1) \;\ldots\; \mathbf{r}_i(P+Q_i)\,]$ of the combined AR stationary model in each block can be estimated using, for example, the covariance method [237, 353] or Bayesian techniques, and are given by $\mathbf{r}_i = \{\mathbf{r}_{\mathbf{a}}, \mathbf{r}_{\mathbf{b}_i}\}$, where $\mathbf{r}_{\mathbf{a}} \equiv [\,\mathbf{r}_{\mathbf{a}}(1) \;\ldots\; \mathbf{r}_{\mathbf{a}}(P)\,]$ are the roots of the IIR filter, $\mathscr{A}$, and $\mathbf{r}_{\mathbf{b}_i} \equiv [\,\mathbf{r}_{\mathbf{b}_i}(1) \;\ldots\; \mathbf{r}_{\mathbf{b}_i}(Q_i)\,]$ are the roots of the TVAR source signal model. In each block, the estimates of these roots are denoted by $\hat{\mathbf{r}}_i = \{\hat{\mathbf{r}}_{\mathbf{a}_i}, \hat{\mathbf{r}}_{\mathbf{b}_i}\}$, where $\hat{\mathbf{r}}_{\mathbf{a}_i}$ is an estimate of $\mathbf{r}_{\mathbf{a}}$ using the data in block i and, correspondingly, $\hat{\mathbf{r}}_{\mathbf{b}_i}$ is an estimate of $\mathbf{r}_{\mathbf{b}_i}$. Estimating the pole locations over each data blocks, $i \in \mathscr{M}$, the following pole sets are obtained:

$$\left\{\begin{array}{l} \hat{\mathbf{r}}_1 = \{\hat{\mathbf{r}}_{\mathbf{a}_1}, \hat{\mathbf{r}}_{\mathbf{b}_1}\} \\ \vdots \;\; = \quad\;\; \vdots \\ \hat{\mathbf{r}}_M = \{\hat{\mathbf{r}}_{\mathbf{a}_M}, \hat{\mathbf{r}}_{\mathbf{b}_M}\} \end{array}\right\} \tag{8.4}$$

Initially, it may appear impossible, without considerable prior knowledge, to partition each pole set, $\hat{\mathbf{r}}_i$, into subsets which separately contain $\hat{\mathbf{r}}_{\mathbf{a}_i}$ and $\hat{\mathbf{r}}_{\mathbf{b}_i}$. However, if all the pole sets $\hat{\mathbf{r}}_i$, $i \in \mathscr{M}$, are compared simultaneously, a number of pole estimates will be contained in a number of *small local regions of support*. It can be hypothesised that each of these subsets contain estimates that are just statistical variations of the same pole. Moreover, most subsets will contain just a few pole estimates, or *elements*, while a few subsets contain many elements: the number of elements within each subset depends on the chosen size of the *local region of support*. Hence, it can be concluded that where the number of estimates within

a particular region is large, the corresponding subset that contains these estimates represents a *stationary* pole, and where the number is small, the subset represents a *nonstationary* pole. If the source signal is modelled as a TVAR process and is known to be comprised only of *nonstationary* poles, and the distorting filter is known to be stationary, the channel can be estimated from some estimate of the *stationary* poles based on the subsets which contain a large number of pole estimates.

This simple histogramme technique can be demonstrated by drawing samples from the probability density function for the AR process described by equation (8.3b) using, for example, the Markov chain Monte Carlo (MCMC) method known as the *Gibbs sampler* [283]. This histogramme technique, while inefficient and lacking robustness, provides insight into the rudimentary estimation mechanism facilitated by the additional degrees of freedom of the BSAR-AR model. Section §8.1.1 provides an overview of how these estimates are sampled, and section §8.1.2 discusses simulation results.

8.1.1 Simulation of Histogram Technique

The histogram method can be demonstrated by obtaining estimates of the AR parameters of the process described by equation (8.3b) in each block. To obtain a large number of estimates, variates from the pdf for this AR process are sampled using the *Gibbs sampler*. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) technique that allows samples to be drawn from complicated probability densities by drawing samples from simpler conditional densities. This sampling method is described in detail in, for example, [283], and the implementation details for this application are given in section §E.1 of Appendix E.

To draw these estimates, suppose the posterior distribution for the BSAR parameters, $\mathbf{b}_i$, and the AR parameters, $\mathbf{a}$, *conditional* on the data in block i, is $p_i \equiv p_i\left(\mathbf{a}_i, \mathbf{b}_i \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}\right)$ where, for clarity, the estimate of the filter parameters, $\mathbf{a}$, conditional on the data in block i, is denoted $\mathbf{a}_i$. The posterior density, p_i, is obtained by writing $c_i(k) = [a_i(p) \star b_i(q)]_k$ and noting that the excitation in block i is related to the observed signal, $x(t)$, by $\mathbf{e}_i = \mathbf{x}_i + \mathbf{X}_i \mathbf{c}_i$, where $[\mathbf{x}_i]_{t-t_i+1} = x(t)$, $t \in \mathscr{T}_i$, $[\mathbf{X}_i]_{t-t_i+1,q} = x(t-q)$, $t \in \mathscr{T}_i, q \in \mathscr{Q}_i$ and $[\mathbf{c}_i]_k = c_i(k)$, $k \in \{1, \ldots, P + Q_i\}$. The Jacobian $J(\mathbf{x}_i, \mathbf{e}_i)$ is unity and, therefore, the likelihood function in data block i is given by:

$$p\left(\mathbf{x}_i \,\middle|\, \mathbf{a}_i,\, \mathbf{b}_i,\, \sigma_i^2,\, \mathbf{x}_{i-1}\right) = \mathcal{N}\left(\mathbf{e}_i \,\middle|\, 0,\, \sigma_i^2\right) \tag{8.5}$$

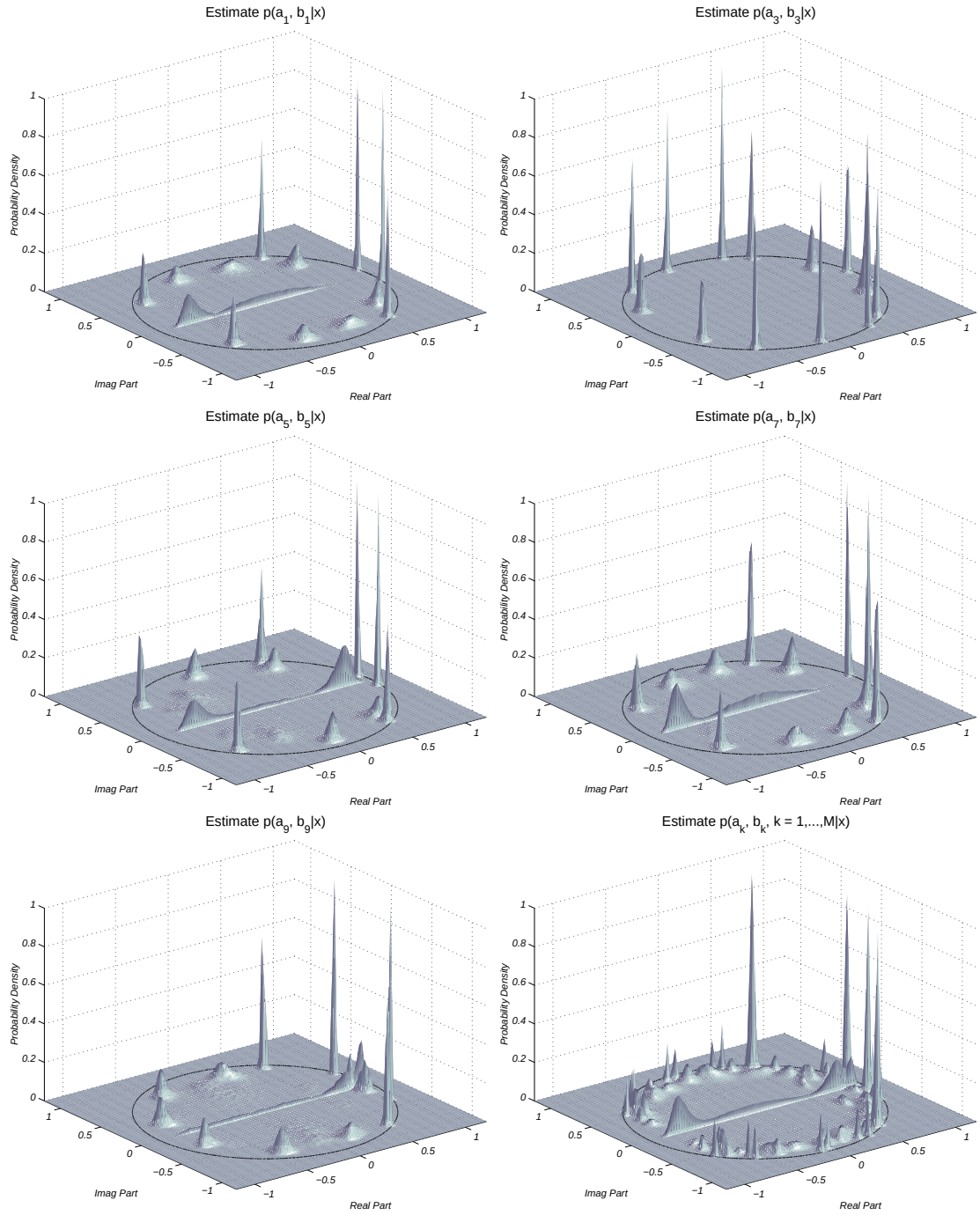


Figure 8.2: *These plots show, from left to right in rows, the histogram over the unit circle of the simulated samples* $\{\mathbf{a}_i, \mathbf{b}_i\}$ *in blocks* $i = \{1, 3, 5, 7, 9\}$ *for the* 2*nd order filter discussed in section* §*8.1.2. The histogram of* all *the simulated samples from* all *the data blocks is also shown in the bottom right hand figure.*

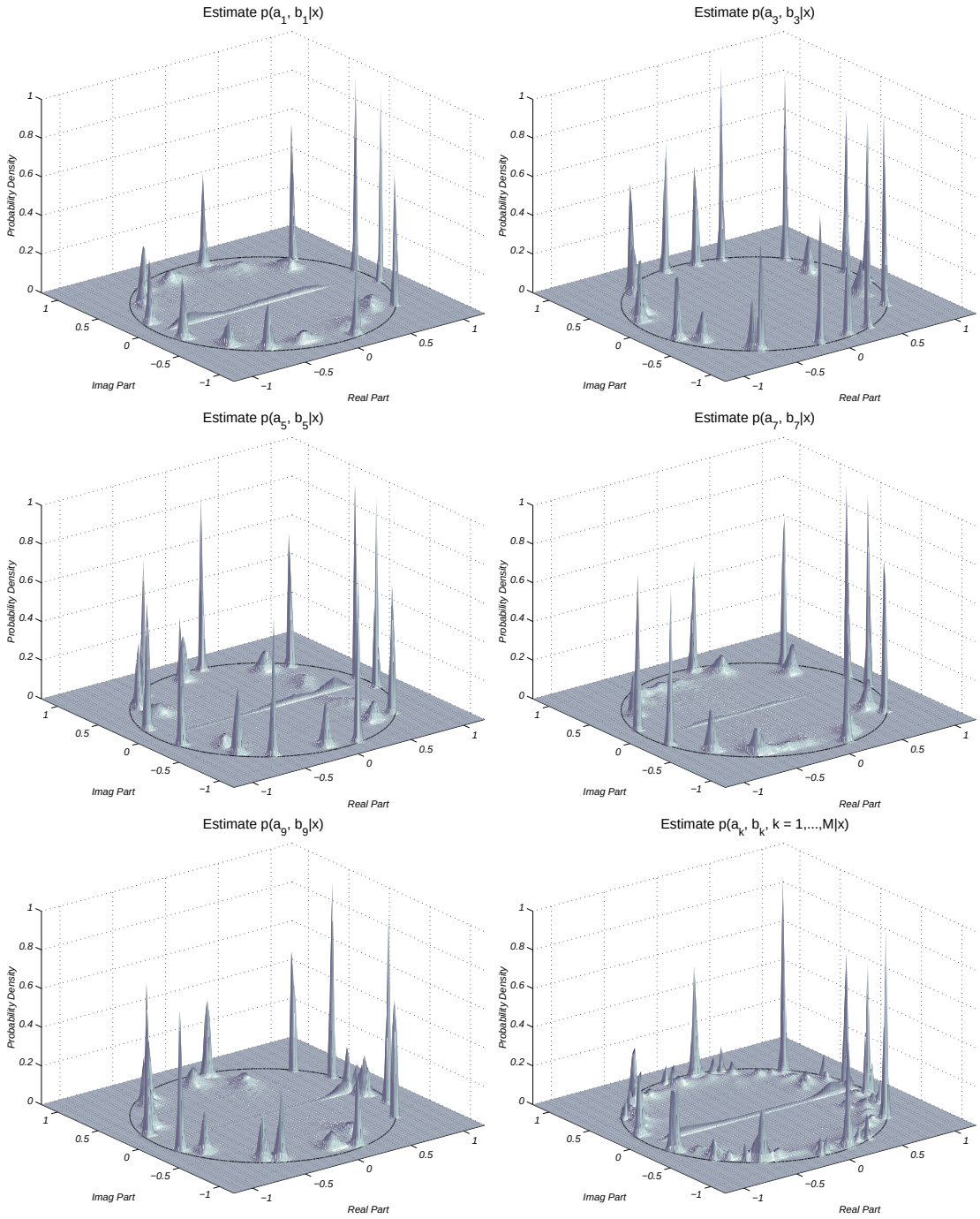


Figure 8.3: *These plots show, from left to right in rows, the histogram over the unit circle of the simulated samples* $\{\mathbf{a}_i, \mathbf{b}_i\}$ *in blocks* $i = \{1, 3, 5, 7, 9\}$ *for the* 8*th-order filter discussed in section* §*8.1.2. The histogram of* all *the simulated samples from* all *the data blocks is also shown in the bottom right hand figure.*

Assuming the prior distributions:

$$\sigma_i^2 \sim \mathcal{IG}\left(\sigma_i^2 \,\middle|\, \frac{\nu_i}{2},\, \frac{\gamma_i}{2}\right)$$
$$\mathbf{c}_i \equiv \begin{bmatrix}\mathbf{a}_i & \mathbf{b}_i\end{bmatrix} \sim \mathcal{N}\left(\mathbf{c}_i \,\middle|\, \mathbf{0}_{P+Q_i},\, \delta_i^2\,\sigma_i^2 \mathbf{I}_{P+Q_i}\right),\ \delta_i \in \mathbb{R}^+,$$

then, using Bayes' rule of equation (7.29b), it follows:

$$p\left(\mathbf{a}_i,\, \mathbf{b}_i,\, \sigma_i^2 \,\middle|\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right) \propto \frac{1}{(\sqrt{2\pi}\sigma_i)^{\hat{R}_i}\,\delta_i^{P+Q_i}} \exp\left\{-\frac{\gamma_i + \mathbf{e}_i^T\,\mathbf{e}_i + \delta_i^{-2}\,\mathbf{c}_i^T\,\mathbf{c}_i}{2\sigma_i^2}\right\} \quad (8.6)$$

where $\hat{R}_i = T_i + P + Q_i + \nu_i + 2$. The excitation variance can be marginalised using the identity in equation (D.5) to give the conditional density $p\left(\mathbf{a}_i,\, \mathbf{b}_i \,\middle|\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right)$. Thus, by sampling and histogramming the variates $\{\mathbf{a}_i,\, \mathbf{b}_i,\, i \in \mathscr{M}\}$ from this distribution, estimates of the parameter $\mathbf{a}$ can be obtained as described in section §8.1.

8.1.2 Simulated Examples

To demonstrate the implicit marginalisation of the parameters of the BSAR source signal model, the 12th-order BSAR synthetic source signal described in section §7.6.1.2 is passed through several different LTI IIR filters, the output of which is observed. The frequency responses of these filters are those shown in Figure 7.10. The model orders of both the filter and the source signal are assumed to be known. The first filter used in the simulation is second-order, and its pole lies at $(r, \theta) = (0.95, \pm 0.3\pi)$: this corresponds to the filter response shown in Figure 7.10(a). The second filter is 8th-order and is shown in Figure 7.10(b): the pole locations of each of these filters are shown in Figure 7.11.

Samples are drawn from the posterior distribution given by equation (8.6) using the Gibbs sampler, as discussed above. In Figures 8.2 and 8.3, the samples are histogrammed on a grid covering the unit circle for each of the filters. The figures show histograms of samples for individual blocks, where it is seen that peaks are located at the positions of the BSAR poles, as well as the positions of the poles due to the filter. However, when the samples from all the data blocks are histogrammed together, it is seen that the peaks are now located at the position of the stationary poles. These stationary poles belong to the filter and to any poles representing the stationary components of the speech signal. The peaks, corresponding to stationary poles, which reside at around 3.7 kHz lie very close indeed to the location of the filter's poles, while the peak that resides at few hundred Hertz corresponds to the

pole representing the pitch of the speech. The pole due to the speech pitch is not quite stationary and, in fact, moves around slightly, as can be seen by looking at the resonant peak near DC in Figure 7.7(b). A blind deconvolution algorithm for a specific application can be designed to ignore this slowly varying pole, provided there is sufficient prior knowledge about the existence of such a pole.

8.2 Bayesian Blind Deconvolution

The histogram technique does not constrain the distorting filter, $\mathscr{A}$, to be stationary across data block boundaries. It is by virtue of the fact the filter is *actually* stationary that the technique can detect the filter parameters by considering the system poles. If the filter really is stationary, a more robust method of parameter estimation results by taking account of this additional prior information. This knowledge is readily incorporated into the Bayesian framework. The Bayesian formulation is easily obtained from the analysis presented in section §7.7.3 since, as discussed in section §7.6, the source signal, $s(t)$, is *modelled* as block stationary. In this model, the output of the system, $x(t)$, is related to the input, $s(t) \equiv s(t, \boldsymbol{\theta}, \mathbf{x})$, by:

$$s(t) = x(t) + \sum_{p=1}^{P} a(p)\, x(t-p) \tag{8.1}$$

The parameters of this system model, $\boldsymbol{\theta}$, is given by $\boldsymbol{\theta} = \{\mathbf{a}, P\}$, where $\mathbf{a} \in \mathbb{R}^P$. The Jacobian in equation (7.34), $\mathscr{J}(\mathbf{x}, \mathbf{s})$, is unity since $\mathbf{x}$ and $\mathbf{s}$ have a linear causal relationship and, therefore, equation (7.39a) reduces to:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) \propto p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi}, \mathcal{I}\right) \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T S_i \left(S_i^T S_i + \delta_i^{-2} I_{Q_i}\right)^{-1} S_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|S_i^T S_i + \delta_i^{-2} I_{Q_i}\right|^{\frac{1}{2}}} \tag{8.7}$$

where $R_i = \frac{\nu_j + T_j + 1}{2}$, $i \in \mathscr{M}$. Equation (8.7) is written in terms of $\mathbf{s}(t)$ to emphasise that the posterior can be efficiently calculated by 'inverse filtering' the data, $\mathbf{x}(t)$, before performing matrix products. A maximum marginal *a posteriori* (MMAP) estimate is used for the parameters $\mathbf{a}$, calculated by finding the global maximum of $p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, P, \mathcal{I}\right)$ using deterministic or stochastic optimisation methods [283,

294]. The accuracy of the MMAP estimate is constrained by the effectiveness of the search algorithm and, as such, by the initial position. It is not the intention of this dissertation to investigate various deterministic or stochastic optimisation techniques, and the *pros* and *cons* of various optimisation methods are left for discussion elsewhere, for example in [283, 294].

8.2.1 Interpretation

It is useful to have an understanding of the mechanism that allows blind deconvolution to be performed using the proposed model. Since the nuisance parameters $\{\mathbf{b}_j, \sigma_j^2, \forall j \in \mathscr{M}\}$, are independent, equation (8.7) may be written in the form:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) = p\left(\boldsymbol{\theta} \,\middle|\, \mathcal{I}\right) \prod_{i=1}^{M} p_i\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, \mathcal{I}\right) \tag{8.8a}$$

where the pdf of the stationary parameters, $\mathbf{a}$, given only the data in the i-th block and the initial conditions, $\mathbf{x}_{i-1}$, is given by:

$$\begin{aligned} p_i &\triangleq p_i\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, \mathcal{I}\right) \\ &\propto \frac{1}{\left|S_i^T S_i + \delta_i^{-2} I_{Q_i}\right|^{\frac{1}{2}}} \left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T S_i \left(S_i^T S_i + \delta_i^{-2} I_{Q_i}\right)^{-1} S_i^T \mathbf{s}_i\right\}^{-R_i} \end{aligned} \tag{8.8b}$$

Considering the case when the TVAR model has fixed model order then, in each block, the observed data, $\mathbf{x}(t)$, is modelled as the result of cascading a stationary AR(Q) model, with parameter $\mathbf{b}_i'$, and a stationary AR(P) model, $P \leq Q$, with parameter $\mathbf{a}'$. Consequently, the pdf p_i has peaks near the *real* Pth-order *factors* of the polynomial with coefficients $\mathbf{a}' \star \mathbf{b}_i'$, since there is no way to distinguish which parts of the parameter sets belongs to the filter and which parts belong to the source signal. This is more readily seen by considering the pole positions of the AR process in each block: since $\mathbf{b}_i'$ is assumed to vary with i, the factors that are 'stationary' are the only ones which remain in the same position, whilst the peaks due to the 'nonstationary' factors will not coincide if the pdfs, p_i, are 'superimposed' over each other. Hence, taking the product of p_i, $\forall i \in \mathscr{M}$, implies that peaks at the stationary points combine constructively, since they coincide, while the peaks at the nonstationary points cancel, since they do not coincide in the parameter space. This is analogous to the argument discussed in section §8.1.

To demonstrate this with an example, consider filtering the BSAR(2) synthetic

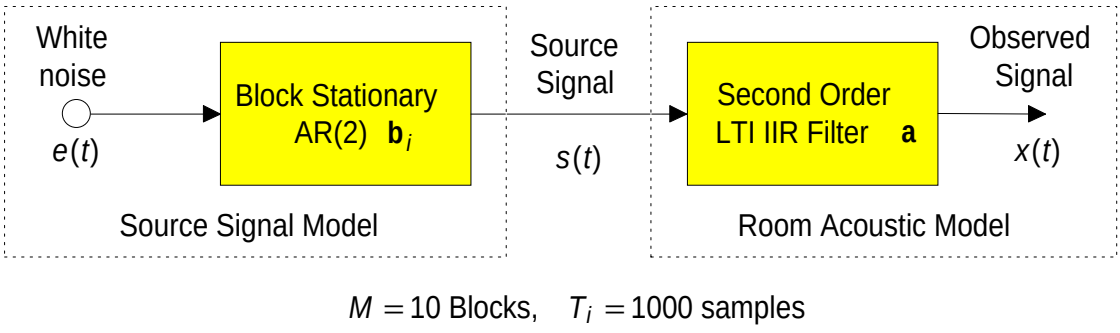
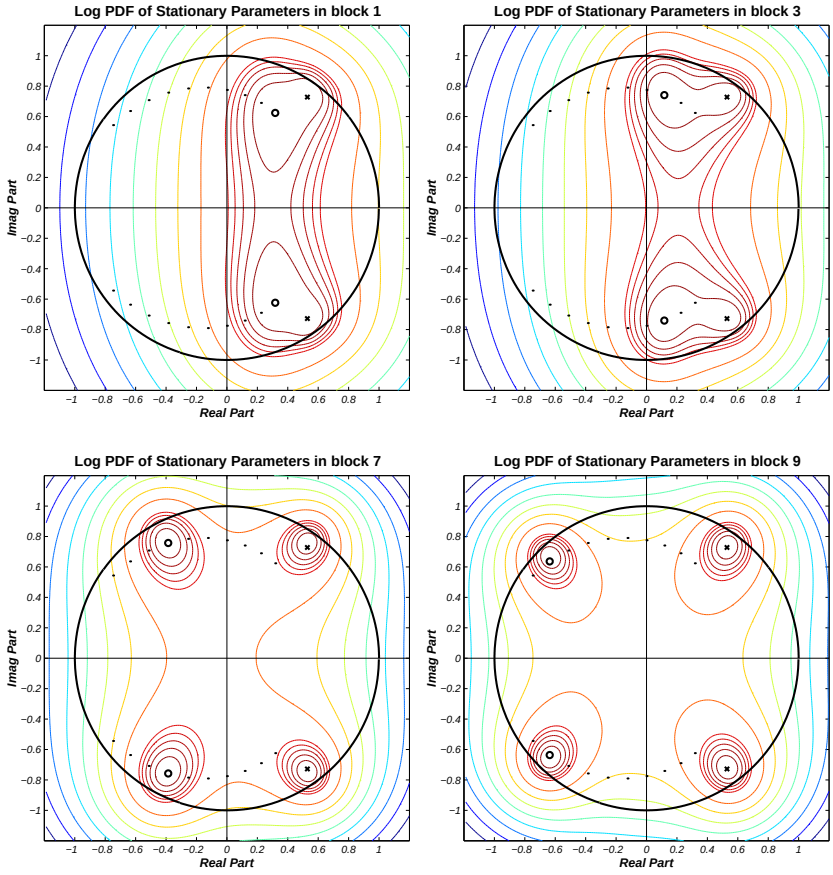


Figure 8.4: *Model used in the investigation of the* maginalisation *process.*

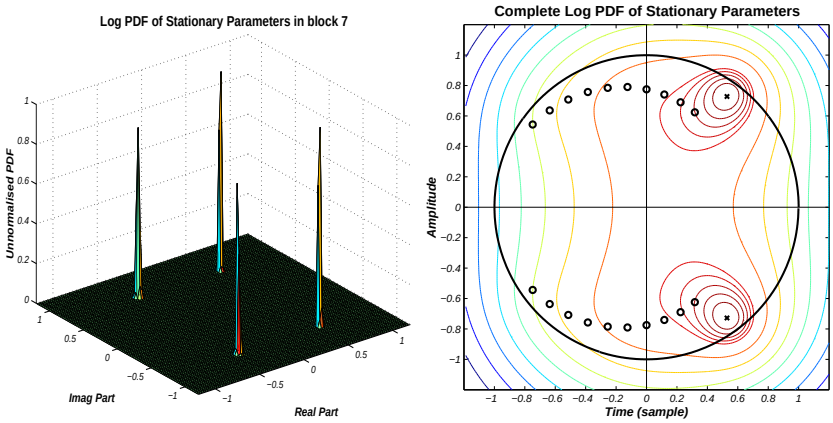
source signal proposed in section §7.6.1.1, denoted $\mathbf{s}(t)$, by a second-order IIR filter, as shown in Figure 8.4. Let the stationary component have conjugate poles at $(r, \theta) = (0.9, \pm 0.3\pi)$, and let the TVAR have conjugate poles which move linearly in radius and phase between the points $(0.75, \pm 0.35\pi)$ and $(0.95, \pm 0.85\pi)$. Further, let $M = 10$ and $T_i = 1000$, $\forall i \in \mathscr{M}$. These pole trajectories are described in equation (7.14), and shown in Figure 7.7(a) on page 175. For a typical data sequence generated by these parameters, Figure 8.5(a) shows the contour plots of $\ln p_i$ for $i = \{1, 3, 7, 9\}$, and Figure 8.5(b) shows a contour plot of $\ln p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \phi, P, \mathcal{I}\right)$ plotted in the complex plane; there is also a plot of p_i for $i = 7$ to highlight the 'sharpness' of the peaks in the distributions. It can be seen that the modes of p_i occur at $\mathbf{a}'$ and $\mathbf{b}_i'$ and, as such, one mode is stationary, whilst the other moves. The estimate of $\mathbf{a}$ is very accurate, as shown by the ratio of the spectra of the actual filter response to the estimated filter response (not plotted for brevity).

8.2.2 Effect of Length and Number of Blocks

The *Cramer-Rao Lower Bound* [286, 353] may be informally stated as *the product of a generally increasing function of block length* T_i *and the variance of a stationary unbiased estimator for the parameters* $\mathbf{a}$*, is greater or equal than some lower bound.* This emphasises an inherent problem in the modelling of nonstationary processes by a block stationary process. If the block length is large, the variance of the parameter estimate is small; however, in that block, the actual parameter values may have changed significantly, such that the model no longer accurately reflects the time-varying nature of the underlying signal. If, on the other hand, the block length is small, the variance of the estimate is large, although the block stationary model will better represent the time-varying nature of the signal.



(a) Plot of $\ln \mathrm{p}_i\left(\mathbf{a} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, \mathrm{P}_i, \mathcal{I}\right)$ for $i = \{1, 3, 7, 9\}$, showing the stationary peak at $\mathbf{a}'$ and the time-varying peak at $\mathbf{b}'_i$.



(b) [Left] $\ln \mathrm{p}\left(\mathbf{a} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, \mathrm{P}, \mathcal{I}\right)$. [Right] $\mathrm{p}_7\left(\mathbf{a} \,\middle|\, \mathbf{x}_7, \mathbf{x}_6, \boldsymbol{\phi}, \mathrm{P}_7, \mathcal{I}\right)$.

Figure 8.5: *The contours are plotted at* $\{10\%, 20\%, \cdots, 90\%, 92\%, 94\%, \cdots, 98\%\}$*, and the small black circle denotes the position of the time-varying pole in the block. The unit circle is also plotted for convenience.*

This raises the question of whether an optimum block length exists. Furthermore, the probability density function of $\mathbf{a}$ is conditional on the block lengths T_i, *i.e.* $p\left(\mathbf{a} \middle| \mathbf{x}, Q_i, t_i, i \in \mathscr{M}, P, \mathcal{I}\right)$ and, thus, T_i could be marginalised numerically. However, for a good understanding of modelling a time-varying system by a block stationary system, the philosophical consequences of this question must be answered, as well as applying standard numerical techniques. The effect of block length when the distorted signal is observed in additive noise is considered in section §8.4.3.

8.3 Effect of Model Order on Parameter Estimation

In section §8.2.2 the effect of the block length on the accuracy of the channel parameter estimate was considered. In the discussions so far, it has also been assumed that the correct form of source signal and channel models are known. This raises two fundamental questions regarding the Bayesian methodology presented so far in this chapter. First, does the form of the model for the source signal and channel accurately represent the system under consideration? Second, if the model is of the correct form, which *model order* is appropriate? Together these questions form a problem referred to as *model selection*. The former question can be answered heuristically and was discussed in Chapters 6 and 7 when considering room acoustics and speech models, respectively, while the latter question is discussed here.

In this section, the source and channel models are assumed to be autoregressive and IIR respectively; the effect of model order on the estimation of the channels parameters is investigated. It is impossible to do an exhaustive investigation on the effect of model order. As such, a set of experiments are carefully selected to provide enough evidence to propose conclusions regarding the effect of model orders. In all, five different sets of experiments are performed, each with different source and channel models, as outlined below:

Second-order LTI IIR filter The distortion filter is second-order which, as in section §8.2.1, allows the posterior distribution for the filter parameters to be plotted in full. The frequency response of this filter is shown in Figure 7.10(a). The following source signals are used:

1. Synthetic BSAR(2) model, as discussed in section §7.6.1.1,
2. Synthetic BSAR(12) model, as discussed in section §7.6.1.2,
3. Extracts of real speech taken from the vocal version of *Suzanne Vega's 'Tom's Diner'*.

8th-order LTI IIR filter The distortion filter with frequency response shown in Figure 7.10(b), facilitates the investigation of higher model orders. The source signals, discussed in cases 2 and 3 above, are used, although only the results for case 2 will be shown, since case 3 does not yield any additional information, inasmuch as the results in each case are virtually identical.

The results for each of these scenarios are discussed in the following subsections.

8.3.1 Second-Order LTI IIR filter

If the distorting filter is *known* to be second-order, then $p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, P, \mathcal{I}\right)$ can be plotted directly, and the effects of varying $\{Q_i, i \in \mathscr{M}\}$ investigated by visualising the changes in $p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, P, \mathcal{I}\right)$; the setup is similar to that shown in Figure 8.4. This, naturally, is feasible only if Q_i is assumed identical for each data block; $Q_i = Q, i \in \mathscr{M}$. If the source signal is a BSAR(2) process, then the interaction between the source signal parameters, $\mathbf{b}_i$, and filter parameters, $\mathbf{a}$, is easily observable, as demonstrated in section §8.2.1. In practice, the model order required to accurately represent speech is around 12 [237] and, therefore, a synthetic BSAR(12) process is also used for the source signal, as well as segments of a real speech signal. The synthetic signals are outlined in sections §7.6.1.1 and §7.6.1.2. The second-order distortion filter is the same as the filter considered in section §8.2.1.

8.3.1.1 BSAR(2) Linear Variation Synthetic Signal Model

In the first experiment, a BSAR(2) process drives a second-order IIR filter, as discussed above, in a similar setup to that shown in Figure 8.4. Figure 8.6 shows in columns, from top to bottom, for different *beliefs* of source signal model orders:

1. Contour and surface plots of the log-pdf, $\ln p_i\left(\mathbf{a} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, P_i, \mathcal{I}\right)$, of the stationary parameters, $\mathbf{a}$, given only the data in a typical data block.
2. A surface plot of the unnormalised pdf, where the expression for this distribution, $p_i\left(\mathbf{a} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, P_i, \mathcal{I}\right)$, is given in equations (8.8).

3. A surface plot of the pdf for the channel parameters given the entire data set across all data blocks, $p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \phi, P, \mathcal{I}\right)$, as given by the expression in (8.7).

It can be seen from the contour plots of $\ln p_i\left(\mathbf{a} \,\middle|\, \mathbf{x}_i, \mathbf{x}_{i-1}, \phi, P_i, \mathcal{I}\right)$ that for any belief of source model order, its modes are located near the pole positions of both the source signal, and the filter, as discussed in section §8.2.1. The corresponding surface plots of the log-pdf and actual pdf for this data block indicate its level of multimodality, and the plot of the complete pdf of the stationary parameters, given *all the data*, demonstrates how pronounced the overall modes are.

To observe the effect of *over-modelling* the source signal, consider the effect on each of these distributions as the hypothesised source model order, Q, is increased. The location of the poles [1] for the source signal, when modelled with order Q, are plotted using a cross (+) on the contour plot of $\ln p_i(\cdot)$ in Figure 8.6. As the hypothesised model order increases from $Q = 2$ to $Q = 14$, it is seen that $\ln p_i(\cdot)$ flattens out. This is explained by considering the location of the extra poles which are attempting to model the source signal. These extra poles are placed symmetrically around the unit circle, with a radius that increases towards unity as the number of poles increase. This behaviour is explained by observing that the extra poles are attempting to model a flat spectrum, since the minimum necessary number of poles are modelling, perfectly adequately, the resonances peaks.[2] There are additional peaks, albeit not very prominent, in the probability distribution at the location of these extra poles and, therefore, the log-pdf flattens out as the number of additional peaks increases.

[1] These pole locations are estimated by taking the source signal in the corresponding data block and estimating the parameters for a Qth-order AR process using the covariance method [237, 353].

[2] If an AR model has poles located at $z = \{r_p, p \in \{1, \ldots, P\}\}$, the magnitude and phase of the frequency response are given by:

$$\left|A(e^{j\omega})\right| = \frac{G}{\prod_{p=1}^{P} |e^{j\omega} - r_p|}, \qquad \angle A(e^{j\omega}) = \omega P - \sum_{p=1}^{P} \angle\left(e^{j\omega} - r_p\right)$$

The magnitude response is equivalent to the product of the distance between the point $e^{j\omega}$ and the poles r_p. The phase response is equivalent to the sum of the angles between the line joining the point $e^{j\omega}$ to the poles r_p and the real axis. If the additional poles are evenly distributed around the unit circle, then there is a high degree of circular symmetry. Since these additional poles are attempting to model a relatively flat spectrum then, to minimise the gain contribution due to the extra poles, they must either be placed at the origin if the spectrum is completely flat or, if the spectrum is noisy, move towards, but remain inside, the unit circle as their number increases: *i.e.* the high circular symmetry means that phase contributions from 'opposite' poles cancel. This property can be verified by a simple imperical investigation. This pole distribution is also noted in [371].

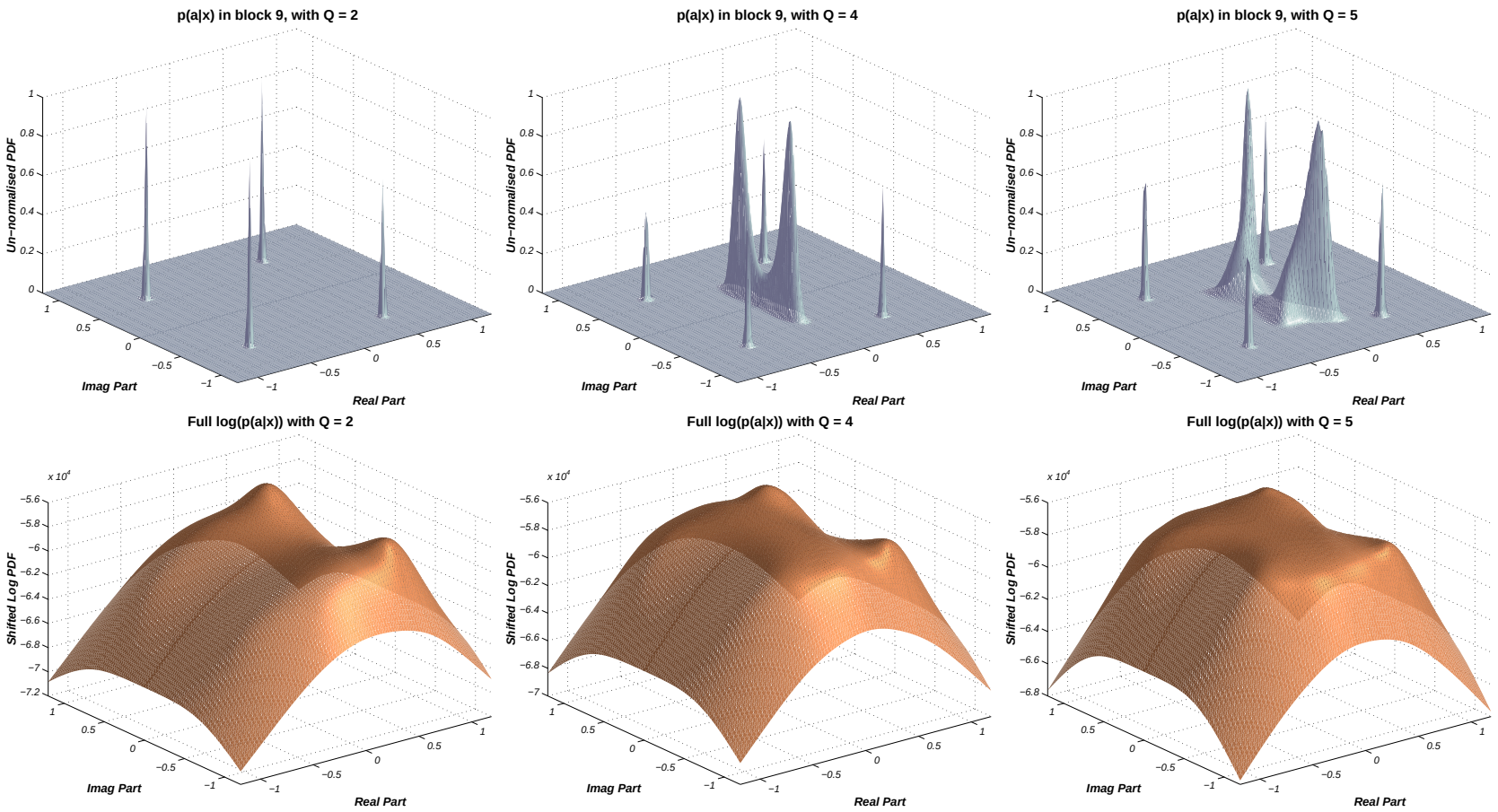
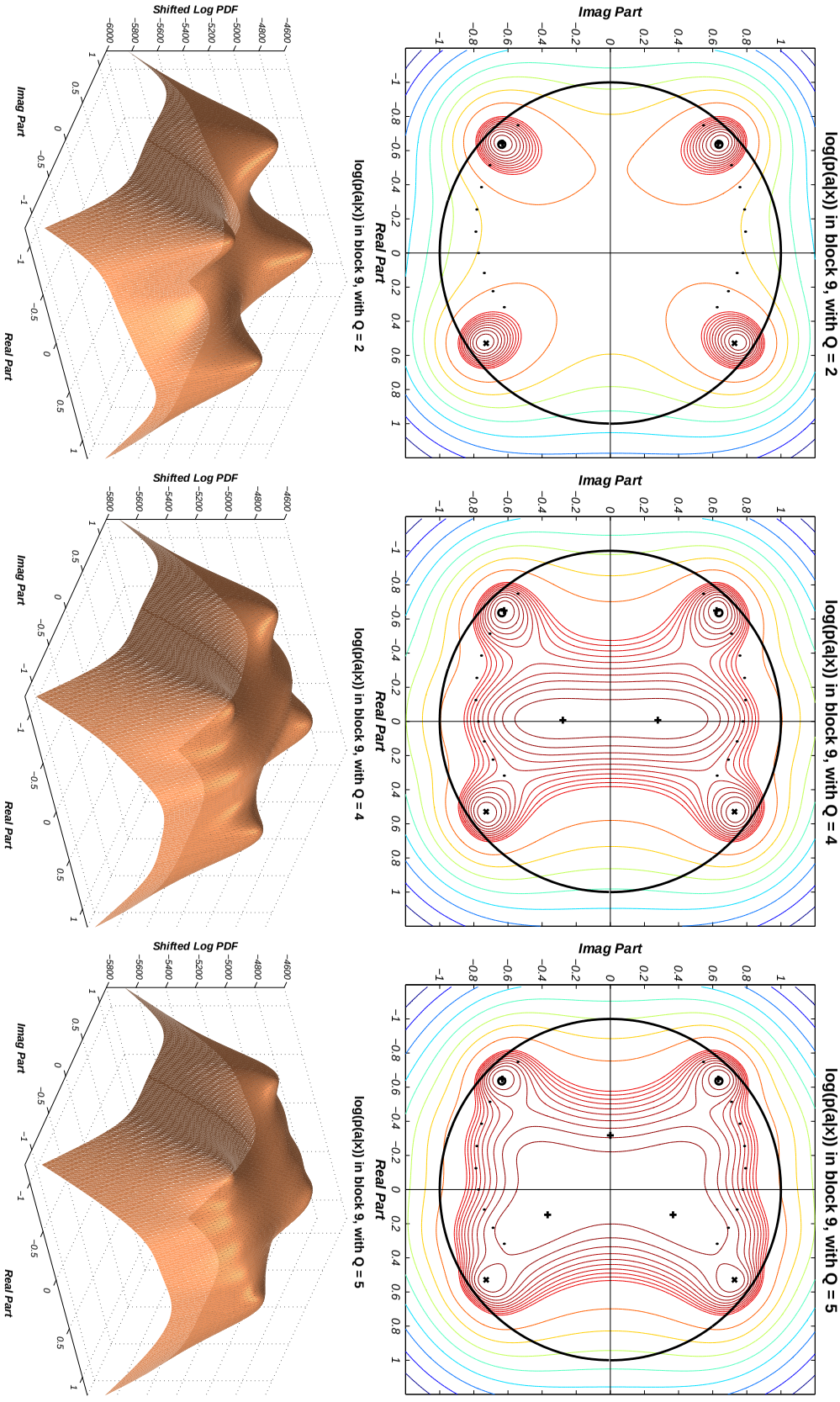


Figure 8.6: *These plots show the effect of model order on parameter estimation and the posterior density for an AR(2) system. From left to right, the TVAR model orders are* $\mathrm{Q} = \{2, 4, 5\}$*. The* ***actual*** *TVAR model order is* 2*, and the block length* $\mathrm{T_i} = 500$*.*


log(p(a|x)) in block 9, with Q = 2
log(p(a|x)) in block 9, with Q = 4
log(p(a|x)) in block 9, with Q = 5
Imag Part
Real Part
Shifted Log PDF

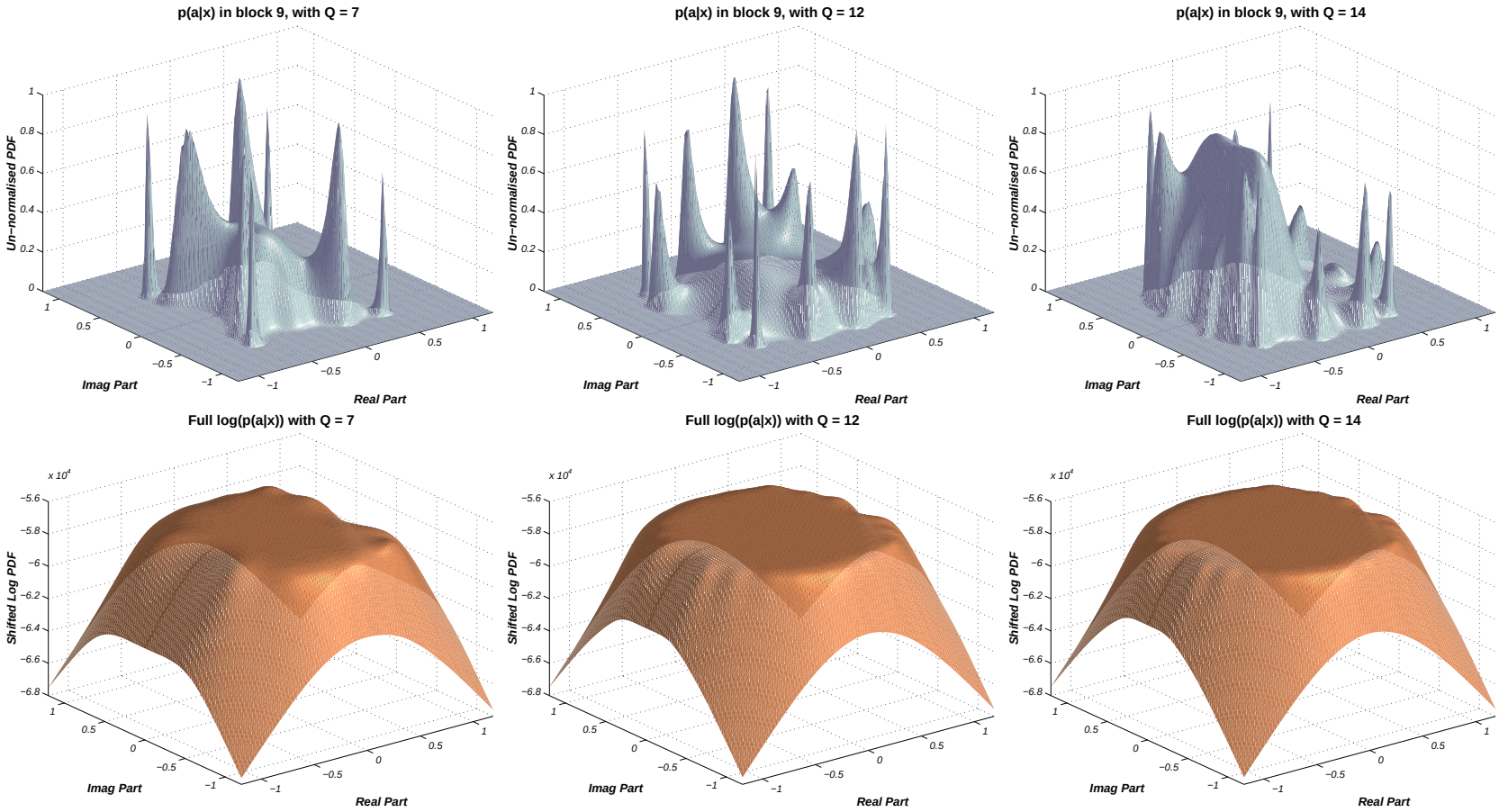
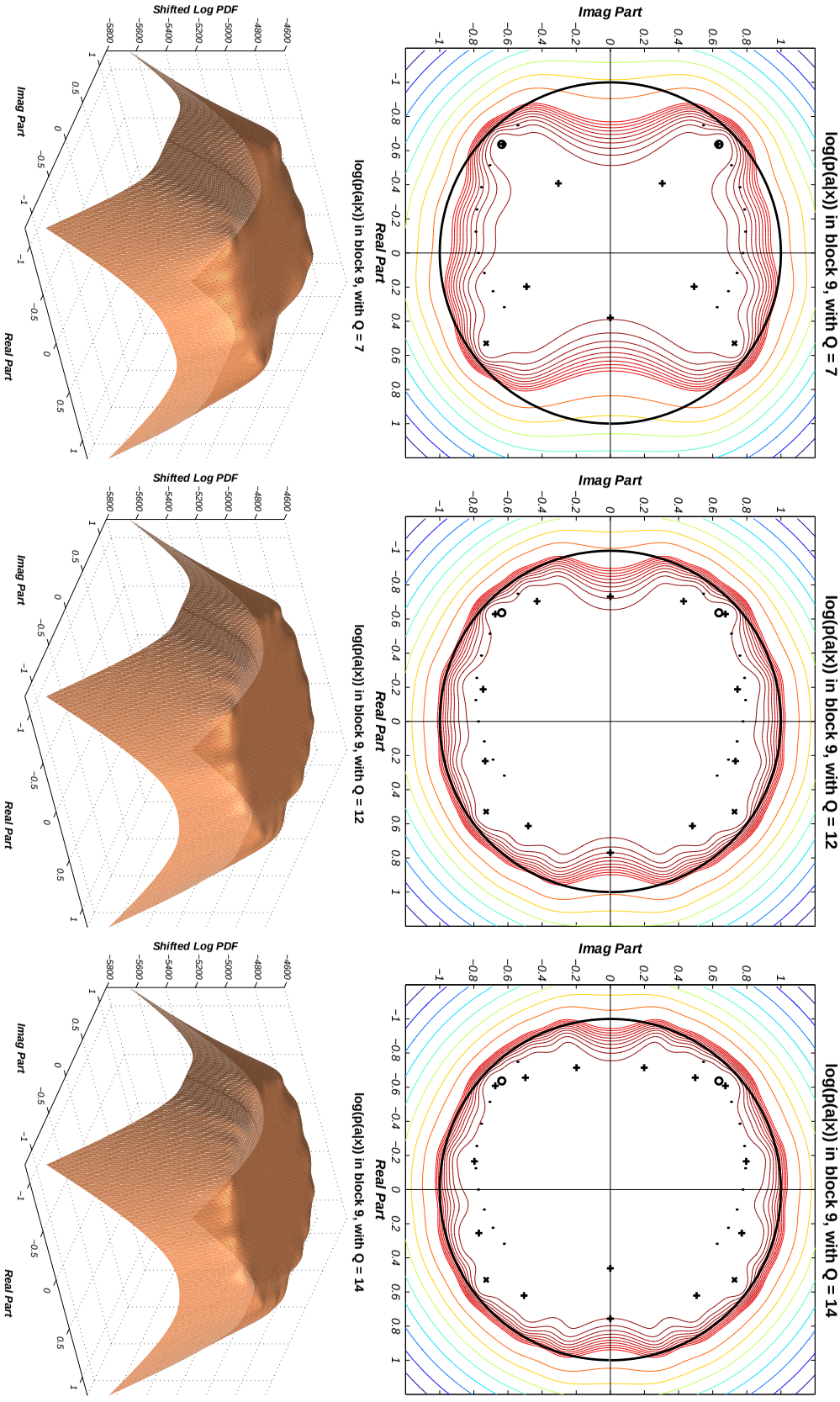


Figure 8.6: *Continued: Left to right, the TVAR model orders are* $\mathrm{Q} = \{7, 12, 14\}$.


log(p(a|x)) in block 9, with Q = 7
log(p(a|x)) in block 9, with Q = 12
log(p(a|x)) in block 9, with Q = 14
Imag Part
Real Part
Shifted Log PDF

However, whilst considerable over-modelling makes it difficult for the peaks corresponding to the resonances of the filter to remain prominent, due to the flattening of the pdf, it is clear that over-modelling the source by a small factor has little impact on the pdf of the channel parameters, given the entire data set. As long as the source signal is nonstationary, the location of these additional poles are time-varying and, as such, despite the individual pdfs for each data block having less pronounced peaks, the peak due to the resonances of the channel remain dominant. The experimental results shown in Figure 8.6 support this argument.

8.3.1.2 BSAR(12) Random Variation Synthetic Signal Model

The second source signal is a realisation of a BSAR(12) process, generated as discussed in section §7.6.1.2. In this case, it is of interest to investigate the effect of the *belief* in source model order on the pdf of the channel parameters *given* the observed data. As seen in Figure 8.7, the effect on $\ln p_i(\cdot)$ is similar to that described in the previous section, where the location of the source signal poles when modelled as Qth-order are shown as large dots $(\cdot)$ on the contour plots; the small dots correspond to the source poles when modelled as 12th-order. The red crosses in the contour plots of $\ln p\left(\mathbf{a} \mid \mathbf{x}, \phi, P, \mathcal{I}\right)$ denote the locations of the resulting channel estimates *if* the entire system is modelled as a 2nd-order stationary system. It is seen that this location corresponds to the pitch resonant of the speech signal as noted in section §7.8. The MMAP estimate obtained using the proposed deconvolution method is shown as a (+), and the actual location of the filter parameters are denoted by an (x).

In this case, it can be seen that *under-modelling* the source signal leads to unsatisfactory results in many cases. If the system is over-modelled, the effects discussed in the previous section are again observed. It is seen in section §8.3.2 that when a resonant, or 'peaky', AR spectrum is under-modelled, the estimated spectrum often results in being relatively flat since the estimator is trying to fit the entire spectrum simultaneously, and not just a particular subband containing one of the resonant peaks. Therefore, when the source signal is under-modelled, the *estimated* source spectrum remains flat, and the pole locations due to the source may appear stationary. The pdf therefore flattens out, and anomalous peaks emerge from these 'false' stationary pole locations.

8.3.1.3 Real Speech as Source Signal

Finally, when the source signal are extracts of real speech, the effect of varying the model order is shown in Figure 8.8. The distributions are plotted for the model order cases corresponding to those used in the BSAR(12) cases in section §8.3.1.2. Since speech is modelled reasonably well by a BSAR process, the results are similar to the synthetic BSAR(12) case. Note that the best channel estimates are obtained when the model order for the source is chosen to be at least 12.

8.3.2 12th-order LTI IIR filter

In the second main model order investigation, a source signal is filtered by the 8th-order LTI IIR filter shown in Figure 7.10(b). For a given belief in source-channel model order (Q, P), a MMAP estimate of $\ln p\left(\mathbf{a} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}, P, \mathcal{I}\right)$ is found using the Nelder-Mead simplex (direct search) method [294] to maximise the logarithm of the expression in equation (8.7).[3] This, therefore, requires the specification of a starting point for the search algorithm. To check whether there is dependance on the starting point, the initial parameter estimates are set as:

$$\mathbf{a}_{\text{ini}} = \mathbf{a}_{\text{act}} \left(1 + \epsilon \ \text{diag}\left[\mathbf{u}\right]\right) \tag{8.9}$$

where $\mathbf{u} \sim \mathcal{N}\left(\mathbf{u} \,\middle|\, 0, 1\right)$, $\mathbf{u} \in \mathbb{R}^P$. The results in Figures 8.9 and 8.10 represent the cases when the initial condition is 'close' to the actual channel parameter vector, corresponding to $\epsilon \ll 1$ (*e.g.* $\epsilon \approx 0.1$), and when the initial condition is 'far' away, corresponding to $\epsilon \gg 1$ (*e.g.* $\epsilon \approx 5$). The dependence on the initial starting location can be removed by using stochastic optimisation methods: for example, MCMC and simulated annealing [283] methods can be applied to ensure the global maximum is always found, independent of the starting position. However, as noted in section §8.2, this is left as further work.

It can be seen from these figures, which plot the estimated channel frequency response and ratio of estimated and actual channel responses, *against* the hypothesised source model order *for* different hypothesised channel model orders,[4] that

[3] In practice, an initial condition may be far from the global maximum, in which case more robust optimisation algorithms are required as discussed in the text. Unless otherwise stated in the text, the *termination tolerance* on the function value is set at 10^{-2}, while the *termination tolerance* on the parameter values is set at 10^{-4}. The maximum number of iterations allowed is set at 5000.

[4] Note that in Figures 8.9 and 8.10, the spectra have been plotted with an offset so that a comparison between individually responses can be made, and to avoid clutter in the images.

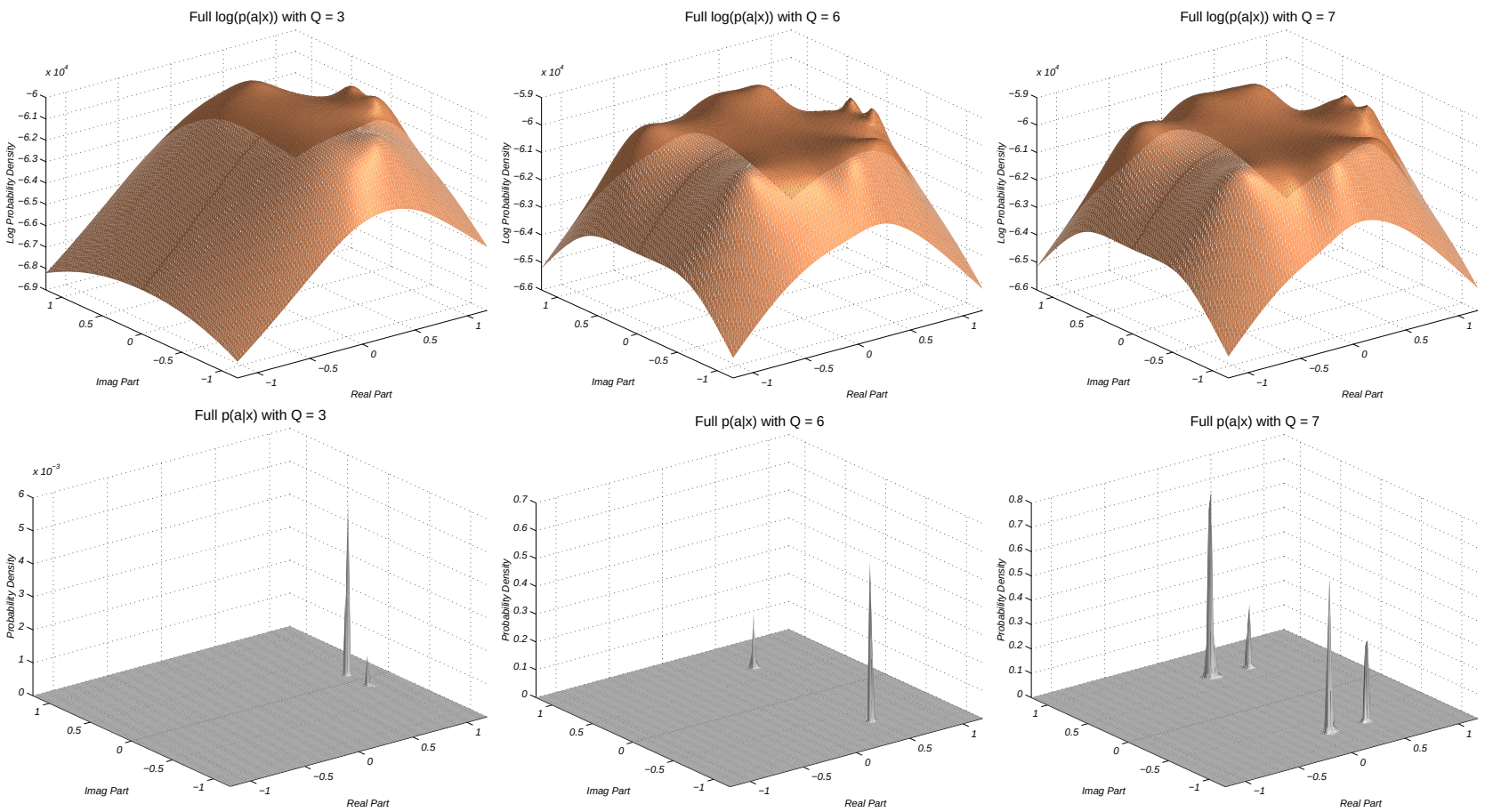
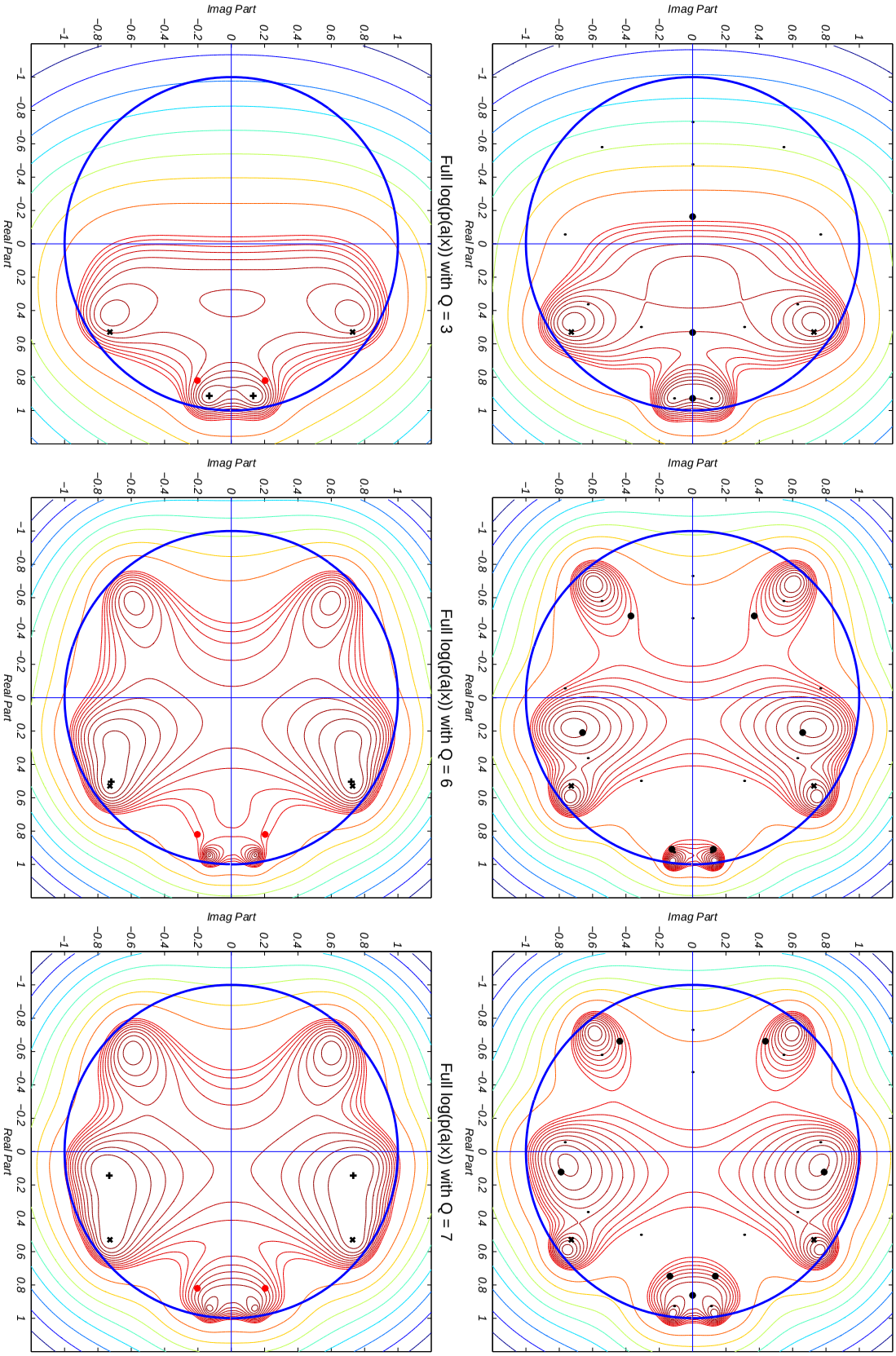


Figure 8.7: *These plots show the effect of model order on parameter estimation and the posterior density for an AR(2) system. From left to right, the TVAR model orders are* $Q = \{3, 6, 7\}$. *The* ***actual*** *TVAR model order is* 12, *and the block length* $T_i = 500$.


log(p(a|x)) in block 5, with Q = 3
log(p(a|x)) in block 5, with Q = 6
log(p(a|x)) in block 5, with Q = 7
Full log(p(a|x)) with Q = 3
Full log(p(a|x)) with Q = 6
Full log(p(a|x)) with Q = 7
Real Part
Imag Part

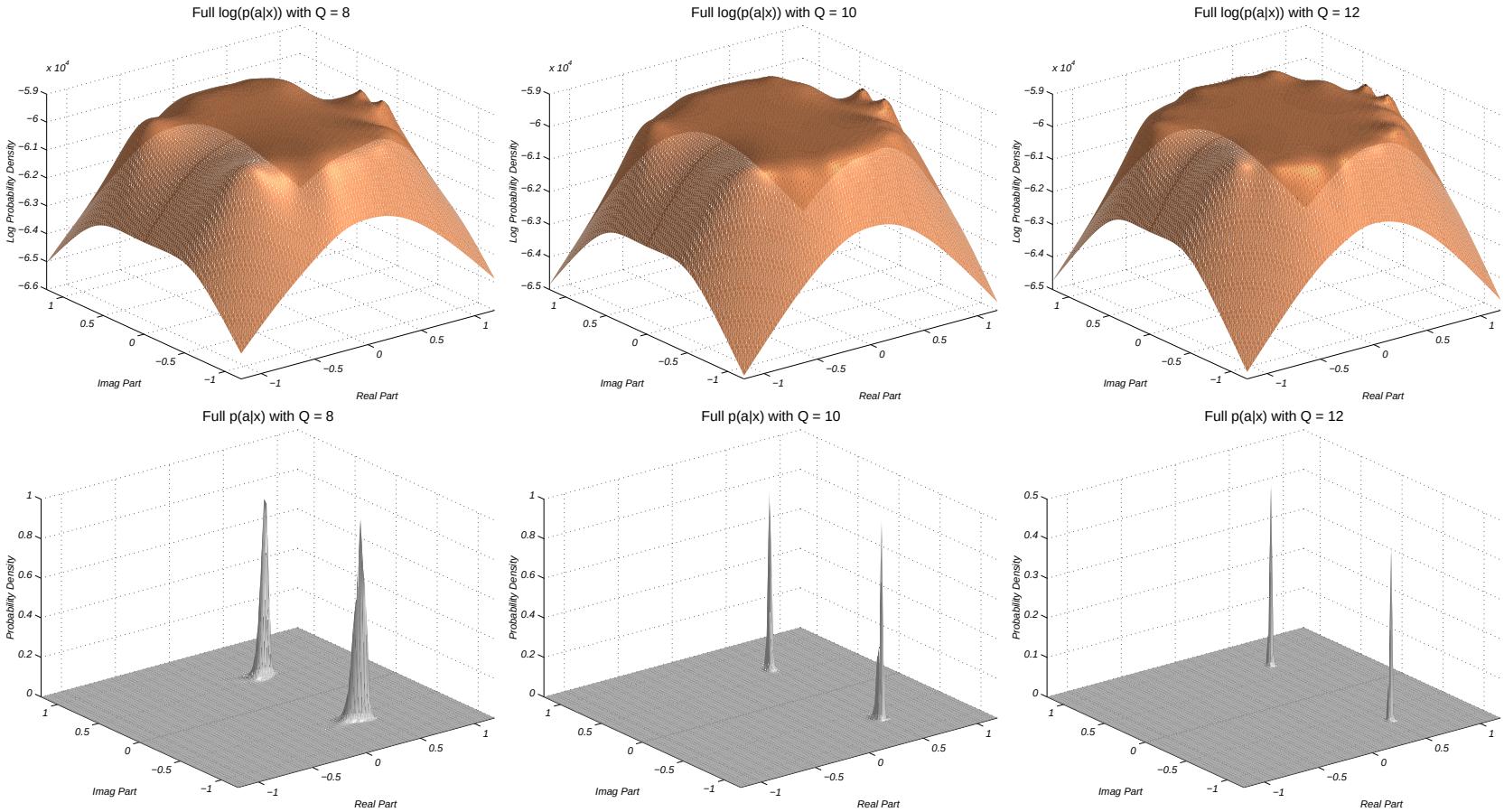
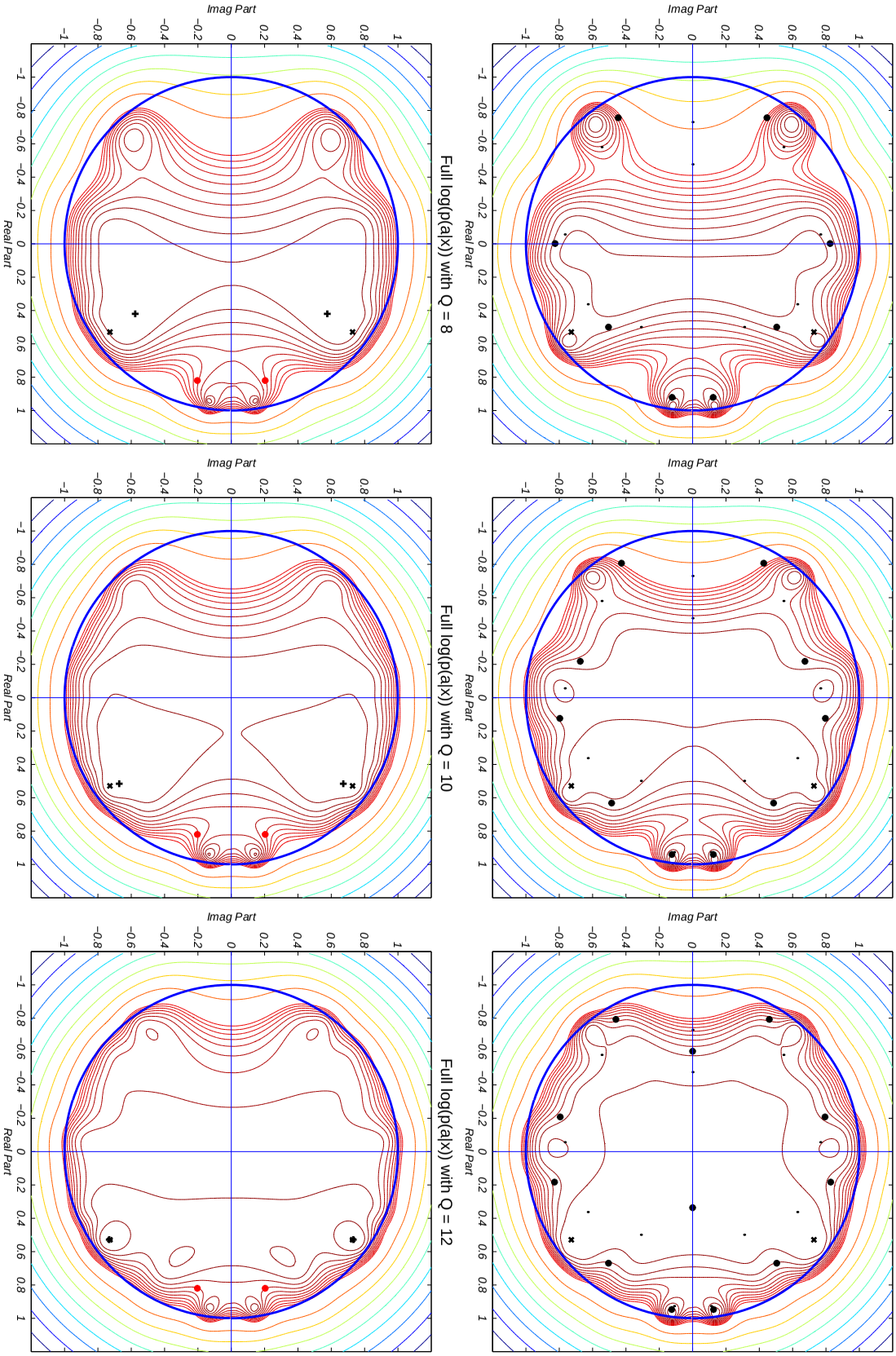


Figure 8.7: *Continued: Left to right, the TVAR model orders are* $\mathrm{Q} = \{8, 10, 12\}$.


log(p(a|x)) in block 5, with Q = 8
log(p(a|x)) in block 5, with Q = 10
log(p(a|x)) in block 5, with Q = 12
Full log(p(a|x)) with Q = 8
Full log(p(a|x)) with Q = 10
Full log(p(a|x)) with Q = 12
Real Part
Imag Part

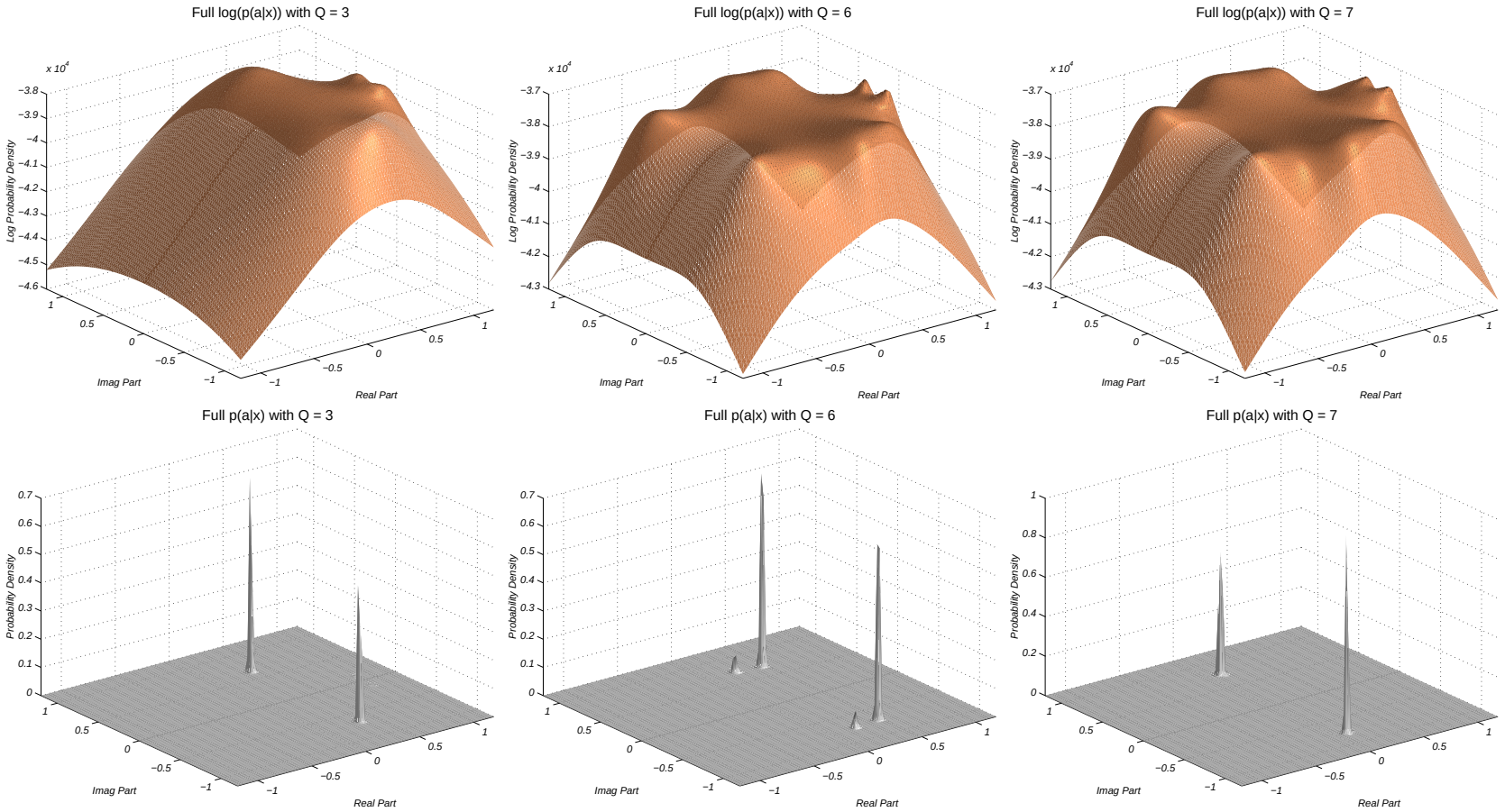
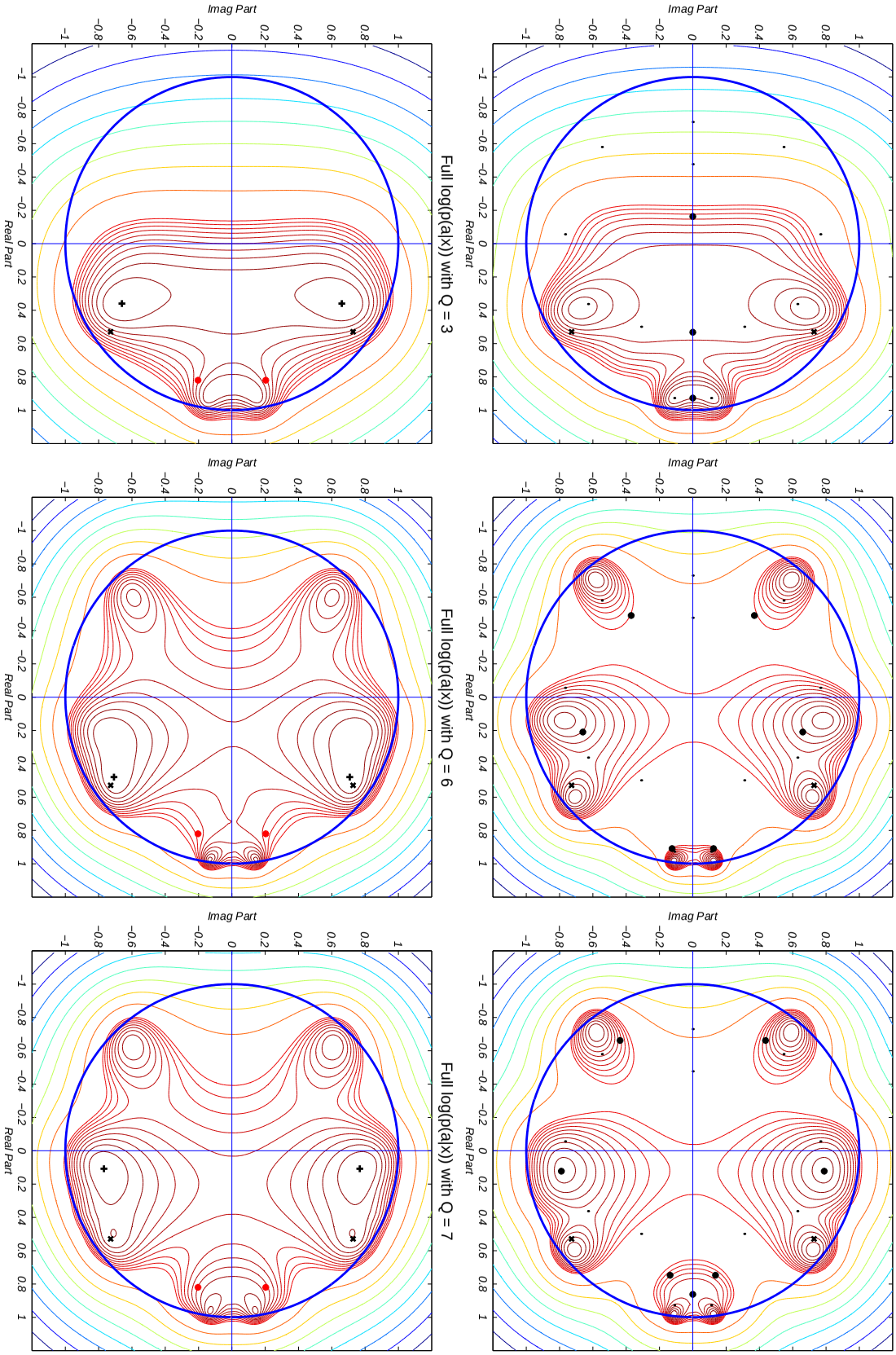


Figure 8.8: *These plots show the effect of model order on parameter estimation and the posterior density for an AR(2) system. From left to right, the TVAR model orders are* $Q = \{3, 6, 7\}$. *The* ***system is driven by real speech,*** *and is modelled as a BSAR(12) process with block length* $T_i = 64$.


log(p(a|x)) in block 5, with Q = 3
log(p(a|x)) in block 5, with Q = 6
log(p(a|x)) in block 5, with Q = 7
Full log(p(a|x)) with Q = 3
Full log(p(a|x)) with Q = 6
Full log(p(a|x)) with Q = 7
Real Part
Imag Part

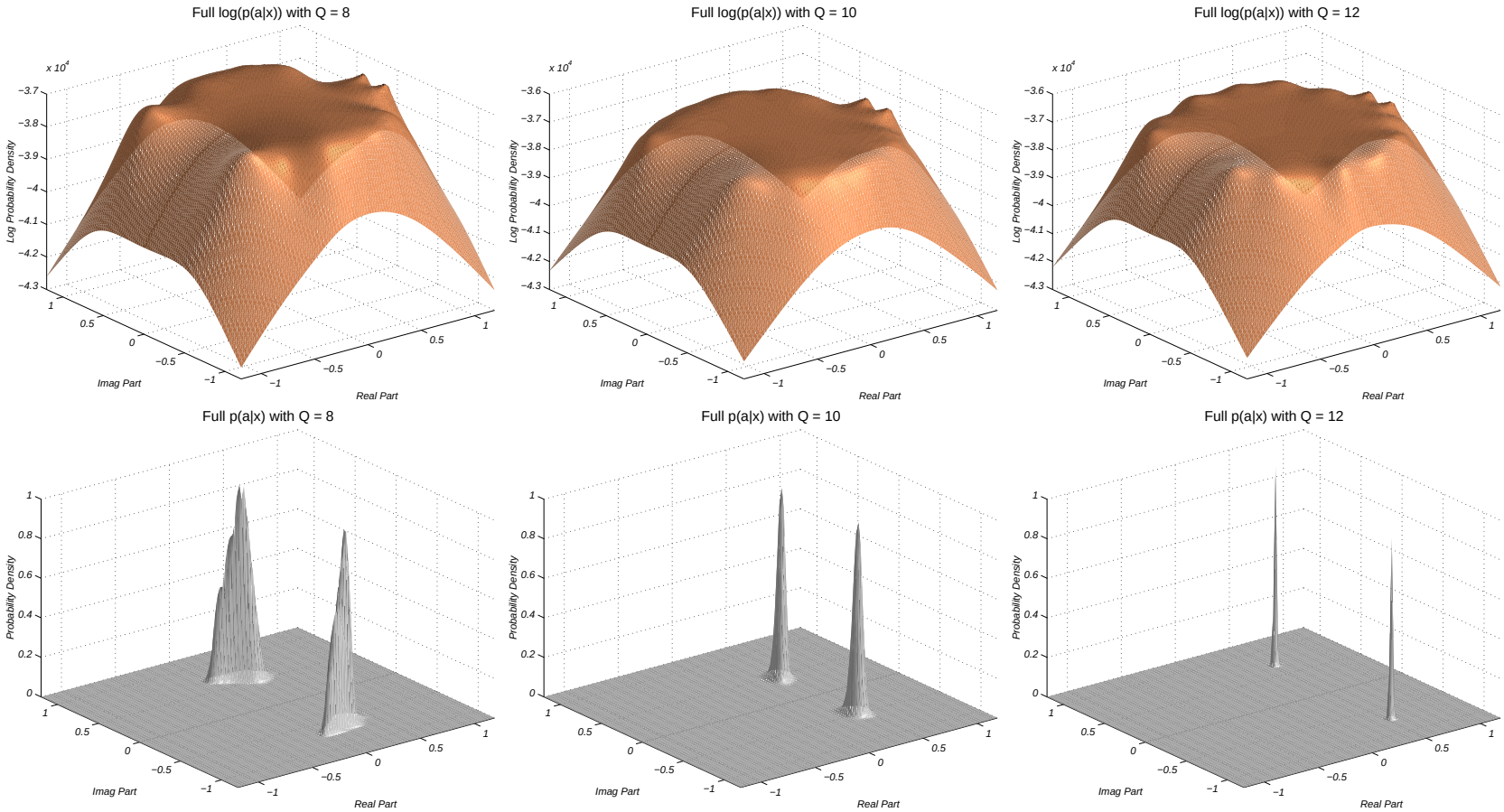
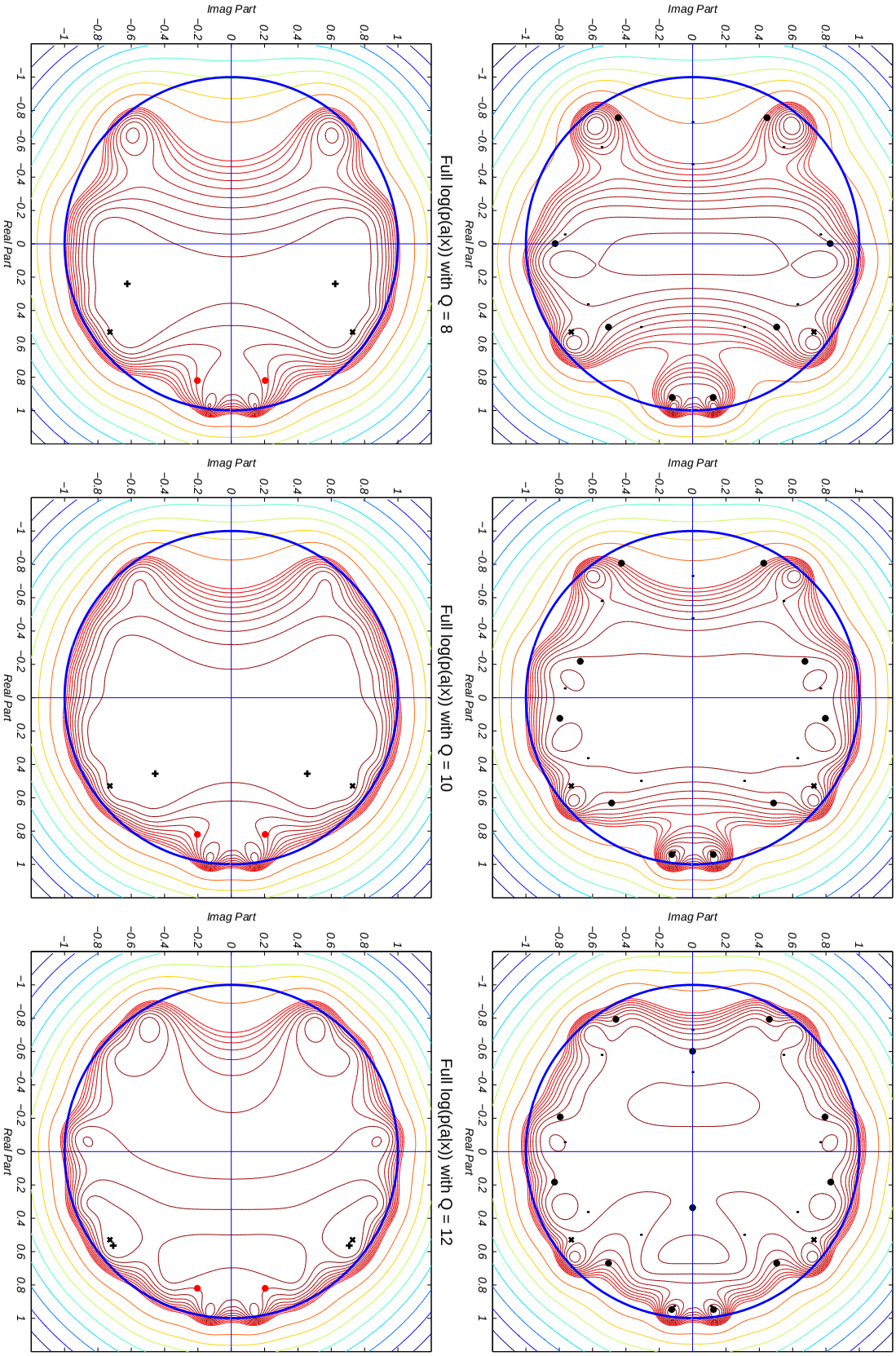


Figure 8.8: *Continued: Left to right, the TVAR model orders are* $\mathrm{Q} = \{8, 10, 12\}$.


log(p(a|x)) in block 5, with Q = 8
log(p(a|x)) in block 5, with Q = 10
log(p(a|x)) in block 5, with Q = 12
Full log(p(a|x)) with Q = 8
Full log(p(a|x)) with Q = 10
Full log(p(a|x)) with Q = 12
Real Part
Imag Part

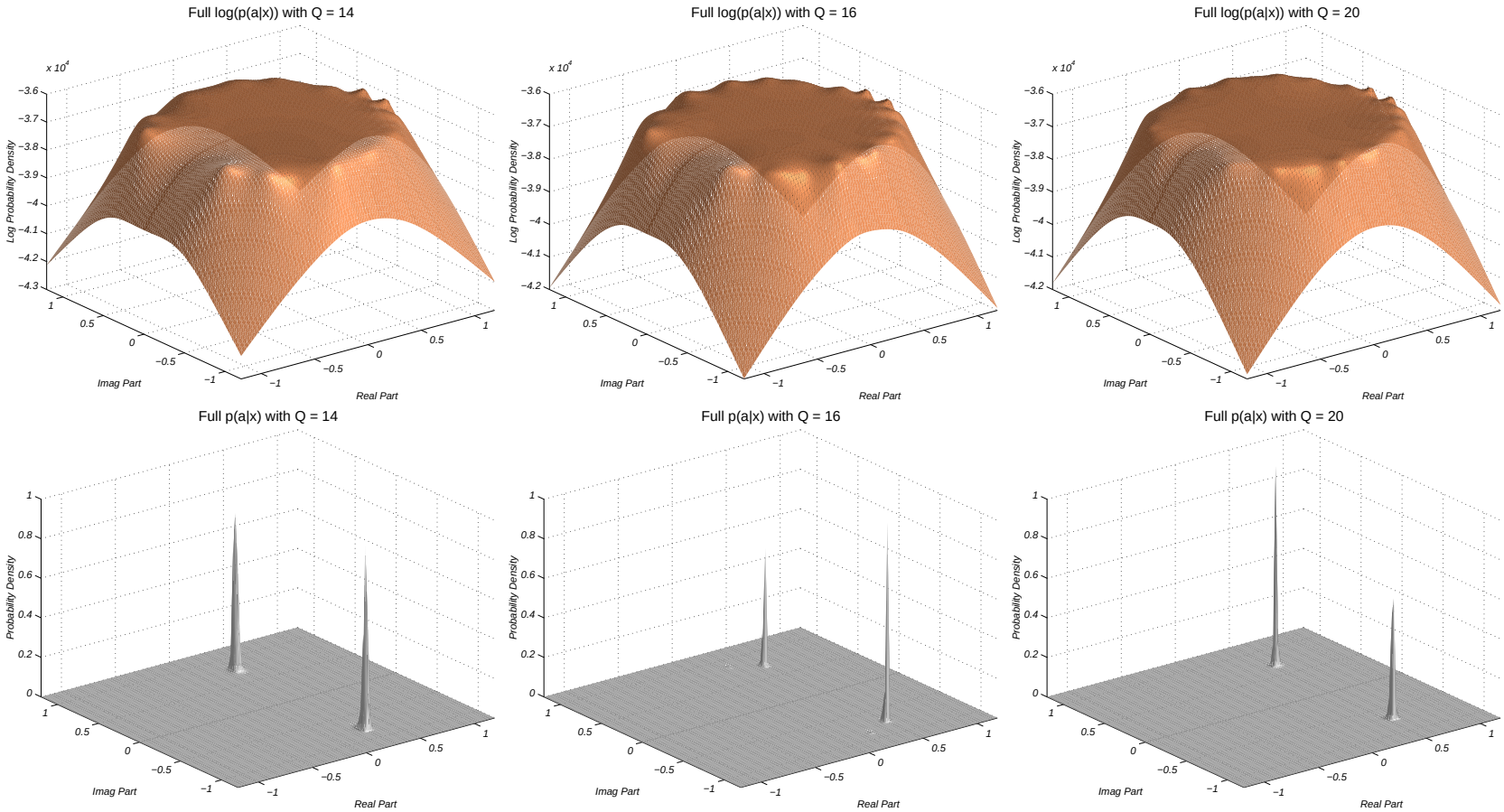
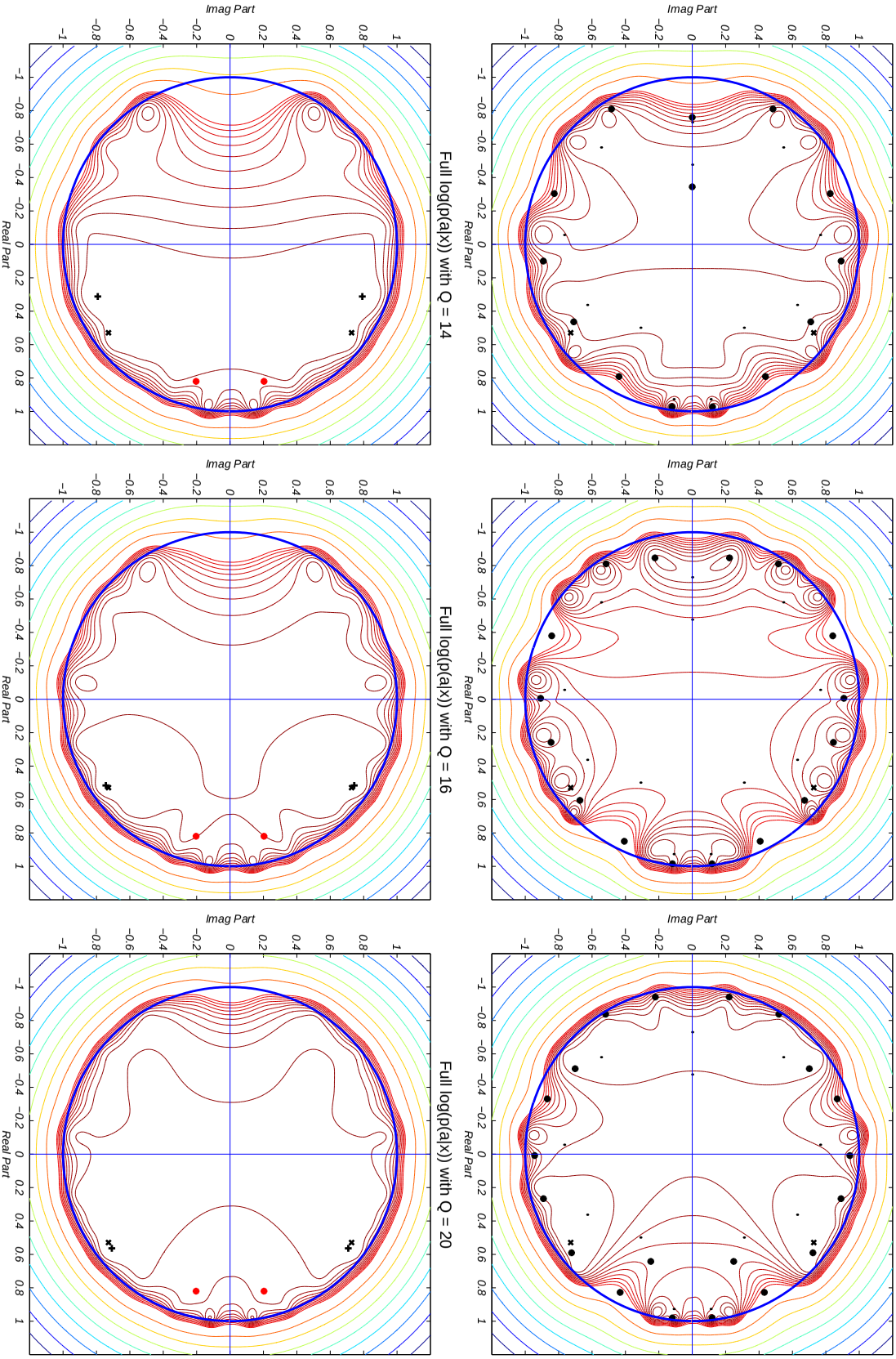


Figure 8.8: *Continued: Left to right, the TVAR model orders are* $\mathrm{Q} = \{14, 16, 20\}$.


log(p(a|x)) in block 5, with Q = 14
log(p(a|x)) in block 5, with Q = 16
log(p(a|x)) in block 5, with Q = 20
Full log(p(a|x)) with Q = 14
Full log(p(a|x)) with Q = 16
Full log(p(a|x)) with Q = 20
Real Part
Imag Part

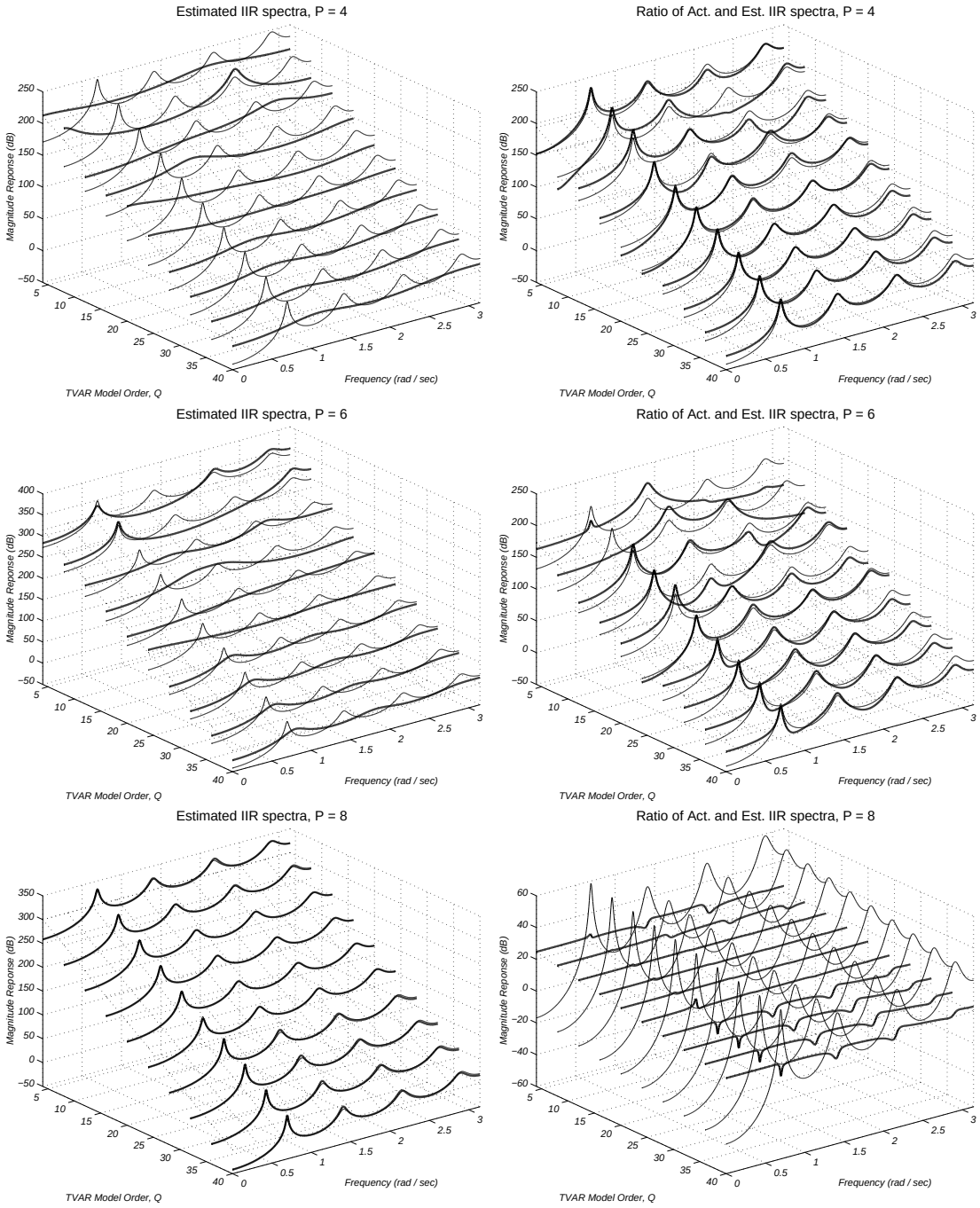


Figure 8.9: *The* (Left Column:) *estimated channel spectrum and* (Right Column:) *ratio of the actual 8th-order channel response to the estimated channel spectrum for* $\mathrm{P} = \{4, 6, 8\}$. *The initial position for the optimisation algorithm is* ***close to*** *the true parameters,* i.e. *in equation* (8.9) $\epsilon \ll 1$, e.g. $\epsilon \approx 0.1$.

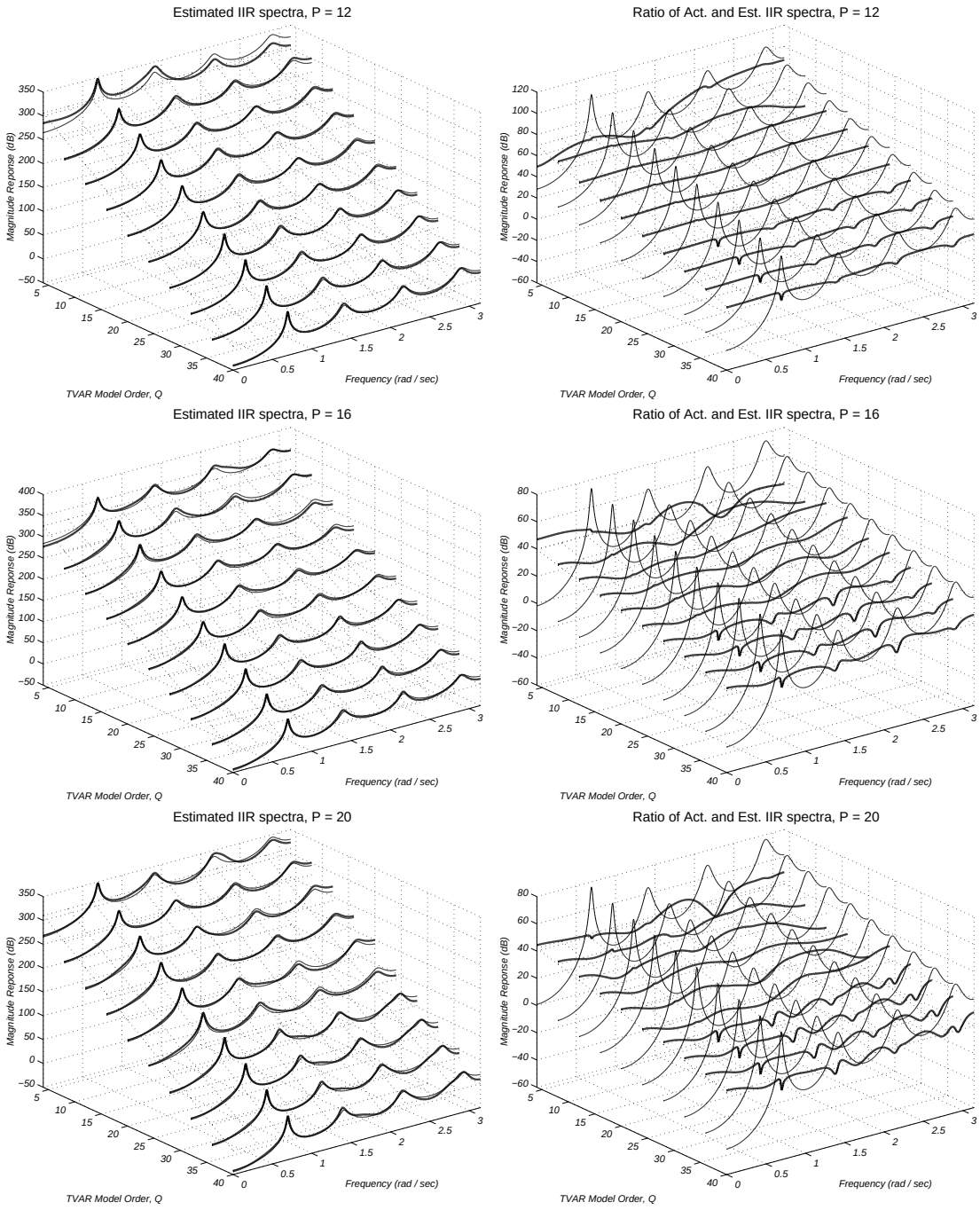


Figure 8.9: *Continued:* Left Column: *Estimated Channel Spectra*. Right Column: *Ratio of actual and estimated channel spectra;* $\mathrm{P} = \{12, 16, 20\}$.

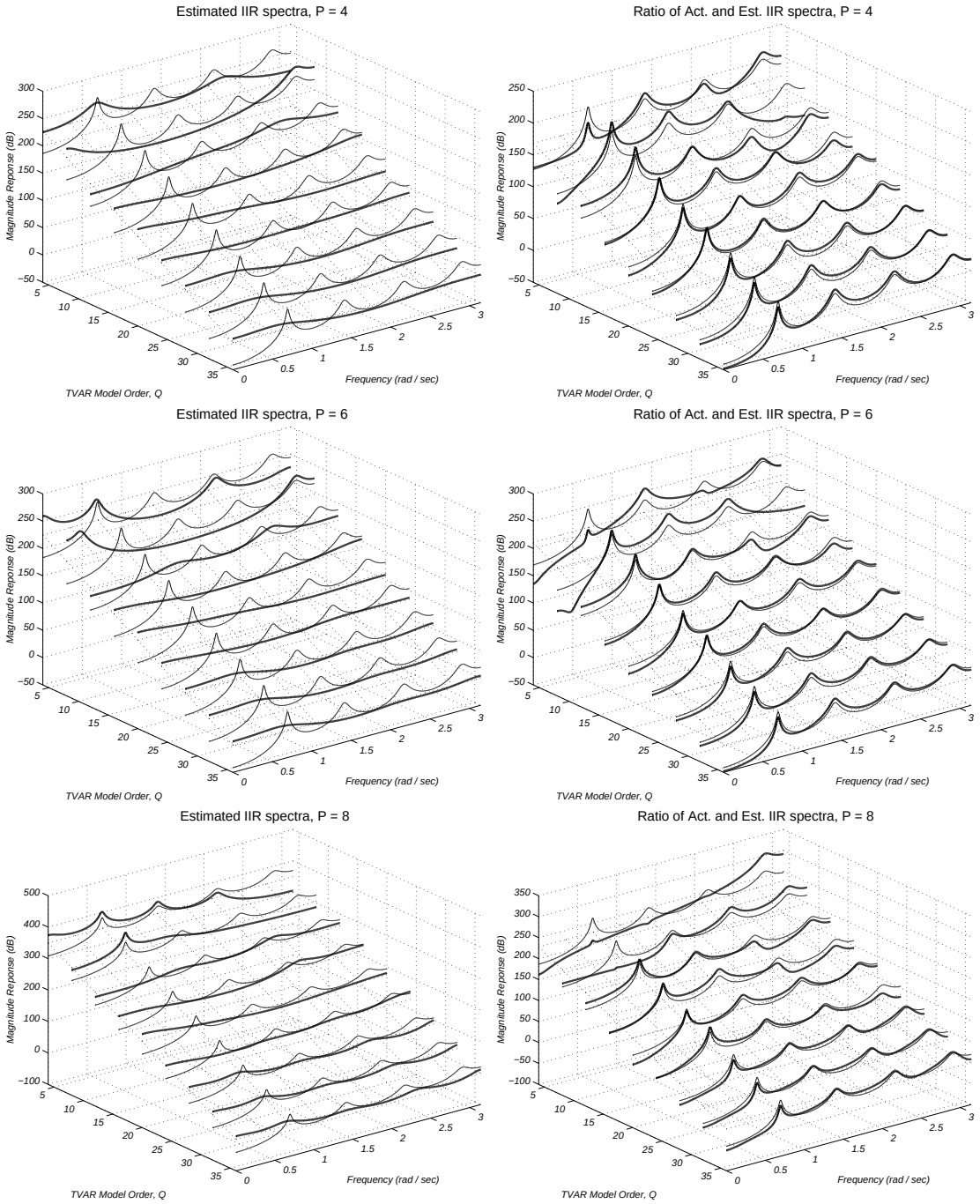


Figure 8.10: *The* (Left Column:) *estimated channel spectrum and* (Right Column:) *ratio of the actual* 8*th-order channel response to the estimated channel spectrum for* $\mathrm{P} = \{4, 6, 8\}$. *The initial position for the optimisation algorithm is* ***far way*** *from the true parameters,* i.e. *in equation* (8.9) $\epsilon \gg 1$, e.g. $\epsilon \approx 5$.

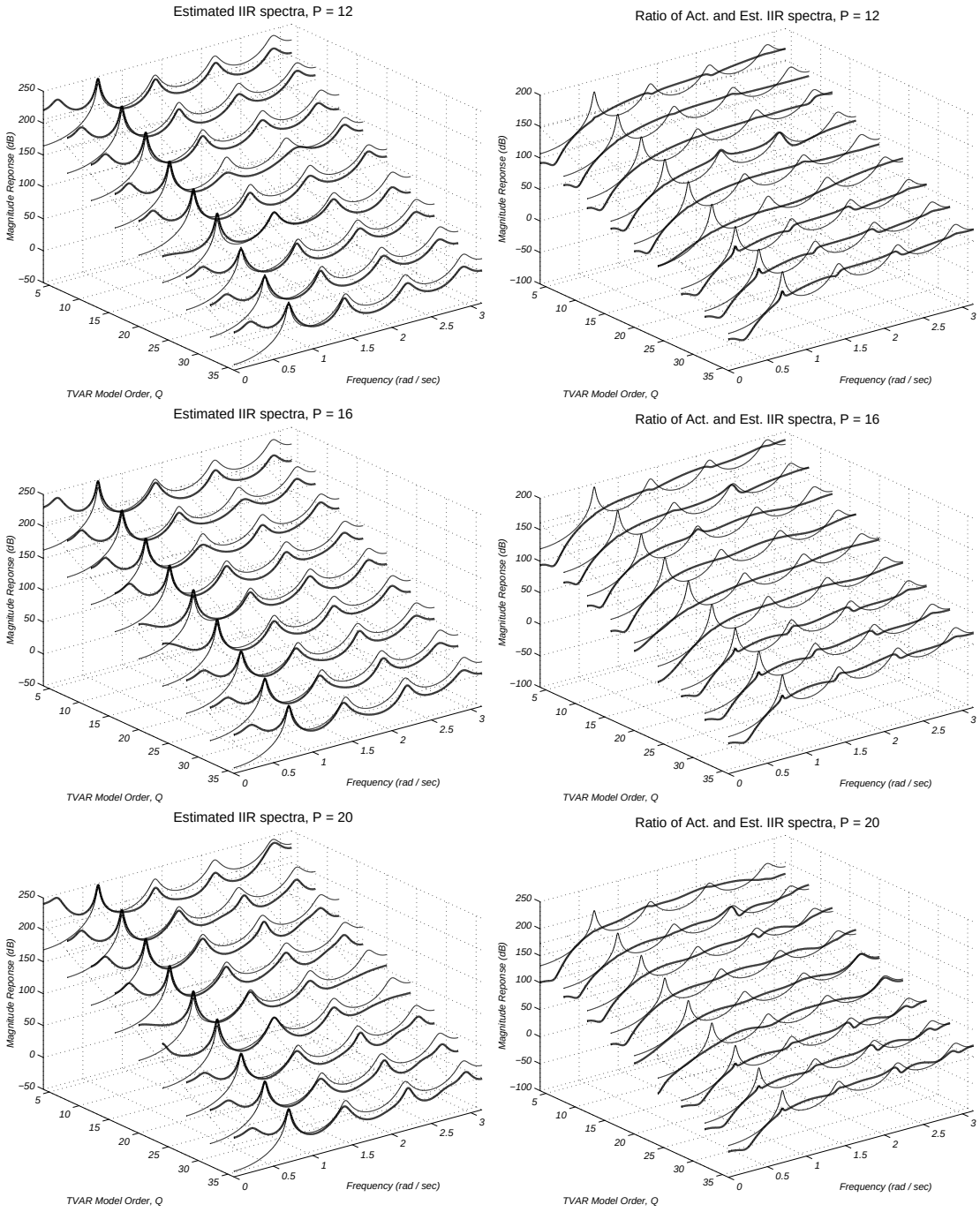


Figure 8.10: *Continued:* Left Column: *Estimated Channel Spectra*. Right Column: *Ratio of actual and estimated channel spectra;* $\mathrm{P} = \{12, 16, 20\}$.

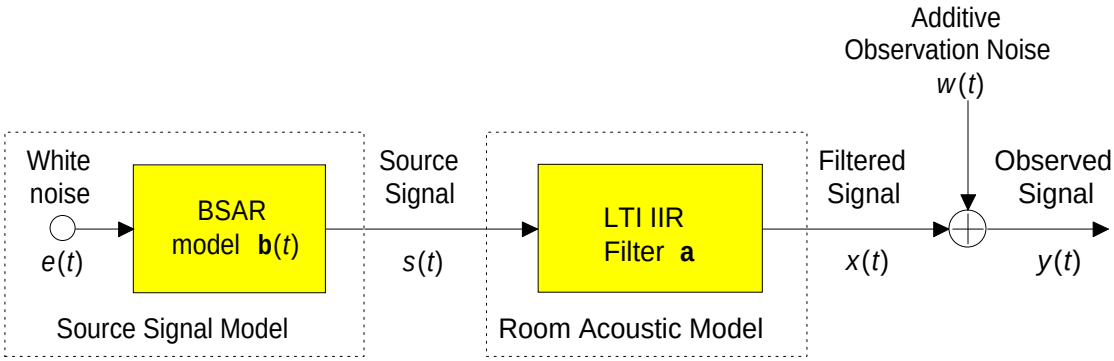


Figure 8.11: *Signal model for the blind deconvolution with observation noise.*

if the *channel* is *over-modelled*, but *not under-modelled*, the estimated spectrum is reasonably accurate and approximately independent of hypothesised source model order. The requirement of over-modelling the channel is simply one of ensuring it is adequately modelled [353]. Under-modelling the source signal is less crucial since the algorithm depends on the nonstationarity of the source signal and not its actual spectrum (see section §8.3.1 for further details). Therefore, if the channel is over-modelled, and the source model order is within a small factor relative to the true model orders, the channel can be accurately estimated.

8.3.3 Nonstationarity vs. Prior Knowledge

The results from the experiments on model order clearly suggest that the utilisation of the nonstationarity in a system reduces the requirement of knowledge regarding the model order. Moreover, although reversible-jump Markov chain Monte Carlo techniques [129] *could* be applied to estimate the model orders, as these sections have shown, accurate model order estimates are not actually required.

8.4 Effect of Observation Noise on Parameter Estimation

The signal model shown in Figure 8.1 does not take account of observation noise. In practice, there exists some observation noise and, as such, a more realistic model,

which takes account of noise, is shown in Figure 8.11. In this model, the observed signal, $y(t)$, is given as the sum of the filtered speech signal, $x(t)$, which is related to the clean speech signal, $s(t)$, by equation (8.1), and the observation noise, $w(t)$, which is WGN with variance σ_w^2, *i.e.* $w(t) \sim \mathcal{N}\left(w(t) \,\middle|\, 0,\, \sigma_w^2\right)$:

$$y(t) = x(t) + w(t) \tag{8.10}$$

As mentioned in section §1.5, the general dereverberation problem can be decomposed into a pure blind deconvolution problem and a pure signal separation problem, the latter which can deal with observation noise. However, in practical applications, it is likely that observation noise will not be completely removed and, consequently, it is of interest to investigate the effect of observation noise on parameter estimation. The next section discusses the influence of SNR on the accuracy of the parameter estimates for a synthetic filter when it is driven by real speech. The effect of noise when the Wiener-Hopf filter (WHF) is used to restore the speech signal is then considered in section §8.4.2. Fo a discussion on other approaches to noise removal see, for example, [120, 378].

8.4.1 Parameter Accuracy vs. SNR

A simple experiment investigating the influence of SNR on the accuracy of the parameter estimates involves driving a synthetic filter with real speech, and calculating the accuracy of the estimates using various cost functions. As before, the parameter estimates are calculated by finding the MMAP estimate of the logarithm of equation (8.7) using the Nelder-Mead simplex method [294]. The initial condition is generated using (8.9) with $\epsilon \approx 0.15$. Generating a number, R, of realisations of the initial condition, $\mathbf{a}_{\text{ini}}^r$, $r \in \mathscr{R} = \{1, \ldots, R\}$, ensures an *average* error function can be determined which is approximately independent of the initial condition.[5] For each of these realisations, the observed signal is generated for a range of SNRs: *i.e.* for each $\mathbf{a}_{\text{ini}}^r$, a realisation of the noise sequence, $\{w(t)\}$, is generated with variance:

$$\sigma_w^2 = \frac{1}{T\, 10^{\frac{SNR}{10}}} \sum_{t \in \mathscr{T}} |x(t)|^2 \tag{8.11}$$

where, as usual, $\mathscr{T} = \{0, \ldots, T-1\}$. There are two measures for the accuracy of the MMAP estimate: the *parameter error function* (PaEF) and the *pole error*

[5] *i.e.* Generating a number of Monte Carlo runs.

function (PoEF). The parameter error function is simply the mean squared error (MSE) between the actual parameters and their estimates, as given by:

$$J_{\mathbf{a}}(\mathrm{SNR}) = \frac{1}{R}\sum_{r=1}^{R} \|\hat{\mathbf{a}}_r(\mathrm{SNR}) - \mathbf{a}\|^2 \tag{8.12}$$

where $\|\cdot\|$ is the Euclidean norm, and $\hat{\mathbf{a}}_r(\mathrm{SNR})$ is the MMAP estimate of $\mathbf{a}$, when the observed process, $y(t)$, has signal-to-noise ratio given by SNR, and the initial condition is given by $\mathbf{a}_{\mathrm{ini}}^{r}$. The pole error function requires a slightly more complicated definition, and is discussed in the next section.

8.4.1.1 Pole Error Function

The pole error function (PoEF) is a measure of the fit of pole estimates to the true pole locations. Denoting the actual pole locations by $\mathbf{r}_{\mathbf{a}} = [\mathbf{r}_{\mathbf{a}}(1) \ \ldots \ \mathbf{r}_{\mathbf{a}}(P)]$, and the estimated pole locations by $\hat{\mathbf{r}}_{\mathbf{a}} = [\hat{\mathbf{r}}_{\mathbf{a}}(1) \ \ldots \ \hat{\mathbf{r}}_{\mathbf{a}}(P)]$,[6] then the PoEF is given by:[7]

$$J_{\mathbf{r}_{\mathbf{a}}} = \min_{\{q(p),\, p\in\mathscr{P}\}} \sum_{p=1}^{P} \|\mathbf{r}_{\mathbf{a}}(p) - \hat{\mathbf{r}}_{\mathbf{a}}(q(p))\|^2 \tag{8.13}$$

where the set $\mathscr{Q}_{\mathrm{perm}} = \{q(p),\, p \in \mathscr{P}\} = \mathrm{perm}\{1,\ldots,P\}$ is a permutation of the elements in $\mathscr{P}$: *i.e.* $J_{\mathbf{r}_{\mathbf{a}}}$ uniquely associates each estimated pole with an actual pole location so as to minimise the total distance between the estimated and actual pole locations. The pole error function can be calculated using Algorithm 8.1 where, for simplicity, it is assumed that P is even, and $\mathbf{r}_{\mathbf{a}}$ consists *only* of complex-conjugate pole pairs such that:[8]

$$\mathbf{r}_{\mathbf{a}} = \left[\mathbf{r}_{\mathbf{a}}(1) \quad \mathbf{r}_{\mathbf{a}}^{*}(1) \quad \mathbf{r}_{\mathbf{a}}(2) \quad \mathbf{r}_{\mathbf{a}}^{*}(2) \quad \ldots \quad \mathbf{r}_{\mathbf{a}}(P/2) \quad \mathbf{r}_{\mathbf{a}}^{*}(P/2)\right] \tag{8.14}$$

where $\Re\{\mathbf{r}_{\mathbf{a}}(p)\} > 0,\ p \in \{1,\ldots,P/2\}$. Further, let $\hat{\mathbf{r}}_{\mathbf{a}}$ be divided into a set of complex-conjugate pairs, and a set of real poles:

$$\hat{\mathbf{r}}_{\mathbf{a}} = \left[\hat{\mathbf{r}}_{\mathbf{a}}(1) \quad \hat{\mathbf{r}}_{\mathbf{a}}^{*}(1) \quad \ldots \quad \hat{\mathbf{r}}_{\mathbf{a}}(P_c) \quad \hat{\mathbf{r}}_{\mathbf{a}}^{*}(P_c) \quad \tilde{\mathbf{r}}_{\mathbf{a}}(1) \quad \ldots \quad \tilde{\mathbf{r}}_{\mathbf{a}}(P_r)\right] \tag{8.15}$$

[6] For clarity, the dependence of the estimated pole locations on the SNR is omitted from the notation in this section.

[7] Note this is defined for a single estimate of the poles. An average pole error function should be obtained from multiple runs to improve accuracy.

[8] A simplification arises since it is known that associating a pole above the real axis with one below the real axis does not lead to the minimum pole error function.

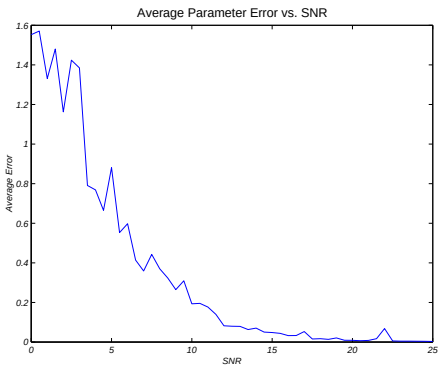
where $\Re\{\hat{\mathbf{r}}_{\mathbf{a}}(p)\} > 0$, $p \in \{1, \ldots, P_c\}$ are the $P_c \leq P/2$ complex poles, and $\{\tilde{\mathbf{r}}_{\mathbf{a}}(p), p \in \{1, \ldots, P_r\}\}$ are the $P_r = P - 2P_c$ real poles, ordered such that $\tilde{\mathbf{r}}_{\mathbf{a}}(p+1) > \tilde{\mathbf{r}}_{\mathbf{a}}(p)$. Let $D_M \in \mathbb{R}^{M \times M}$ be a matrix containing the distances between each of the true pole locations and the estimated pole positions, as constructed at the start of Algorithm 8.1. Further, let D_{M-1}^{-m} be the matrix D_M with the m-th row and first column removed. Then the PoEF, $J_{\mathbf{r}_a}$, defined in (8.13) can be calculated using the recursive algorithm in Algorithm 8.1. For high model orders, the PoEF can be evaluated using the computationally more efficient *out-of-kilter* algorithm [24].

Algorithm 8.1 Calculating the Pole Error Function.

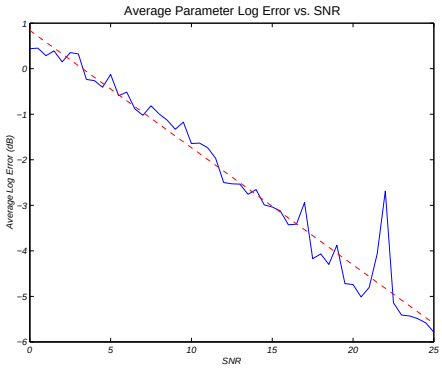
```
for p = 1 to P/2 do
   for q = 1 to P_c do
      [D_P]_{p,q} = 2 ||r̂_a(p) − r_a(q)||
   end for
   for q = 1 to P_r do
      [D_P]_{p,P_c+q} = ||r̃_a(2q − 1) − r_a(p)|| + ||r̃_a(2q) − r_a(p)||
   end for
end for
J_{r_a} = MinimumDistance(D_P)

function d_min = MinimumDistance(D_M)
if M = 1 then
   d_min = D_M
else
   for m = 1 to M do
      d(m) = [D_M]_{m,1} + MinimumDistance(D_{M-1}^{-m})
   end for
   d_min = min(d)
end if
return
```

8.4.1.2 Results

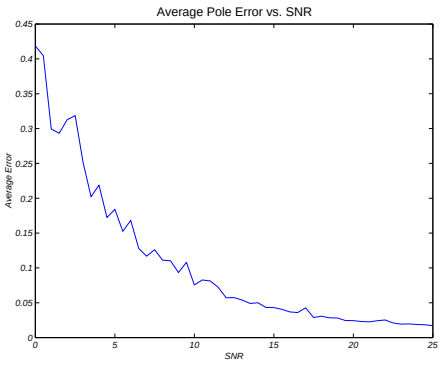
In the simulation results given here, a typical 8th-order IIR filter is driven by speech sampled at 11.025 kHz. With 50 realisations of the initial condition, and SNRs between 0 dB and 25 dB, the results are shown in Figure 8.12(b): Figure 8.12(a) shows the average parameter error function (PaEF), while Figure 8.12(b) shows the log PaEF. The parameter error falls off exponentially with SNR, as reflected by the approximate linear relationship in the plot of the log error. Intererstingly, the PoEF plotted in Figures 8.12(c) and 8.12(d) also falls off exponentially.



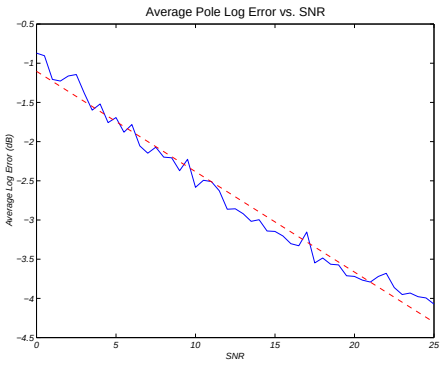
(a) Parameter Error vs. SNR (dB). The relationship is approximately exponential as highlighted in Figure 8.12(b).



(b) Parameter Log Error vs. SNR (dB). The dashed line is the linear regression of the log error using least-squares.



(c) Pole Error vs. SNR (dB). The relationship is approximately exponential as highlighted in Figure 8.12(d).



(d) Pole Log Error vs. SNR (dB). The dashed line is the linear regression of the log error using least-squares.

Figure 8.12: *Effect of observation noise on accuracy of parameter estimates.*

8.4.2 Wiener Filter Restoration

In the last section, it is seen that there is degradation in the accuracy of parameter estimation when there is observation noise. The problem with the existing parameter estimation technique is that the implicit inverse filtering of equation (8.1) used to obtain an estimate of the clean signal, $s(t)$, which is then used in the calculation of equation (8.7), has a noise gain given by (see footnote 25 on page 147):

$$G = 1 + \sum_{p=1}^{P} a^2(p) \geq 1 \tag{8.16}$$

Although the observation noise should be built into the posterior distribution for the parameters $\mathbf{a}$, this would lead to an intractable distribution where numerical methods are required. A simpler, and less rigorous, approach to reducing the effect of the observation noise is to use the Wiener-Hopf filter to obtain an estimate of the clean signal, $s(t)$, rather than a direct inverse, as reflected in equation (8.1). Hence, (8.1) is replaced by:

$$\hat{s}(t) = \sum_{q=0}^{Q} h(t,q)\, y(t-q) \tag{8.17}$$

where the filter, $h(t,q)$, is chosen to minimise the *mean squared error* (MSE), $E\left[\varepsilon^2(t)\right]$, between the estimate of the desired signal of equation (8.17), $\hat{s}(t)$, and the actual desired signal, $s(t)$:

$$E\left[\varepsilon^2(t)\right] \triangleq E\left[|\hat{s}(t) - s(t)|^2\right] \tag{3.1}$$

The resulting filter, $h(t,q)$, is the Wiener-Hopf filter, as introduced in Theorem 1 and section §C.2, [9] and given by the solution of:

$$R_{sy}(t, t-q) = \sum_{\hat{q}=0}^{Q} h(t,\hat{q})\, R_{yy}(t-\hat{q}, t-q), \quad \forall (t,q) \in \mathscr{T} \times \mathscr{Q} \tag{8.18}$$

where $\mathscr{Q} = \{1, \ldots, Q\}$. Since the clean speech, $s(t)$, is nonstationary, then so is the filtered speech, $x(t)$, and, therefore, so is the observed signal $y(t)$. However, $y(t)$ is block stationary and, as such, the correlation functions in (8.18) reduce to:

$$\left.\begin{aligned} R_{sy}(t, t-q) &\equiv R^i_{sy}(q) \\ R_{yy}(t-\hat{q}, t-q) &\equiv R^i_{yy}(q-\hat{q}) \end{aligned}\right\} \quad t \in \mathscr{T}_i,\ i \in \mathscr{M} \tag{8.19a}$$

where $R^i_{sy}(q)$ and $R^i_{yy}(q)$ denote stationary correlation functions in block i, and $\mathscr{T}_i = \{t_i, \ldots, t_{i+1} - 1\} \subset \mathbb{Z}^{T_i}$. Theorem 2 on page 38 states that if $s(t)$ and $d(t)$ are wide sense stationary, then the Wiener-Hopf filter is linear time-invariant. Thus, since these processes are block stationary, it is reasonable to make the approximation that the WHF is LTI over block i:

$$h(t,q) = h_i(q), \qquad q \in \mathscr{Q}_i = \{1, \ldots, Q_i\},\ t \in \mathscr{T}_i,\ i \in \mathscr{M} \tag{8.19b}$$

[9] Note due to that the definition in equation (8.17) is different in form to that in equation (3.4) and, as such, the form of the Wiener-Hopf filter is slightly different to that in Theorem 25.

This assumption is better in the middle of the data blocks, but breaks down near the block boundaries. Nevertheless, the assumption facilitates simple and fast implemententation. Hence, in data block i, equation (8.18) reduces to:

$$R^i_{sy}(q) = \sum_{\hat{q}=0}^{Q_i} h_i(\hat{q})\, R^i_{yy}(q-\hat{q}), \qquad \forall (t,q) \in \mathscr{T}_i \times \mathscr{Q}_i \tag{8.20}$$

which is the familiar form of the stationary Wiener-Hopf filter. From (8.1):

$$R^i_{sy}(q) = R^i_{xy}(q) + \sum_{p=1}^{P} a(p)\, R^i_{xy}(q-p) \tag{8.21}$$

and, using equation (8.10), $R^i_{xy}(q) = R^i_{yy}(q) - R^i_{wy}(q)$. Moreover, assuming $x(t)$ and $w(t)$ are independent, then $R^i_{wy}(q) = R^i_{ww}(q)$ and, therefore, using (8.21), equation (8.20) may be written as:

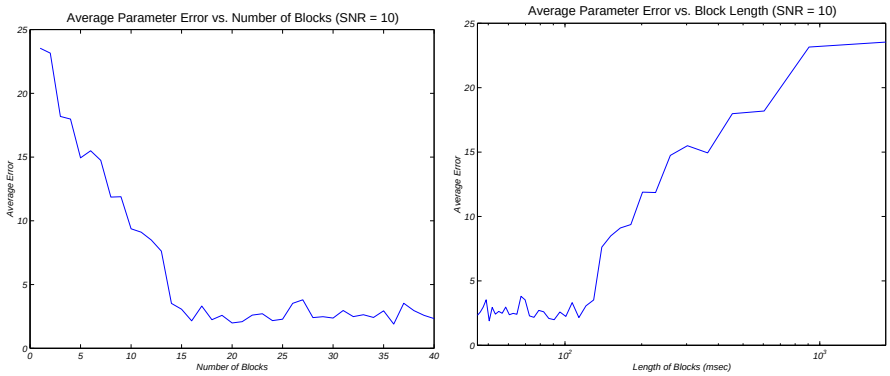
$$\sum_{p=0}^{P} a(p)\left\{R^i_{yy}(q-p) - \sigma_w^2\,\delta(q-p)\right\} = \sum_{\hat{q}=0}^{Q_i} h_i(\hat{q})\, R^i_{yy}(q-\hat{q}),\ \forall (t,q) \in \mathscr{T}_i \times \mathscr{Q}_i$$

where $a(0) \triangleq 1$, and $R^i_{ww}(q) = \sigma_w^2\,\delta(q)$ since $w(t)$ is WGN. This may be written in matrix form as:

$$\begin{bmatrix} R^i_{yy}(0) - \sigma_w^2 & \cdots & R^i_{yy}(P) \\ \vdots & \ddots & \ddots \\ R^i_{yy}(P) & \cdots & R^i_{yy}(0) - \sigma_w^2 \\ R^i_{yy}(P+1) & \cdots & R^i_{yy}(1) \\ \vdots & \ddots & \vdots \\ R^i_{yy}(Q_i) & \cdots & R^i_{yy}(Q_i - P) \end{bmatrix} \begin{bmatrix} a(0) \\ \vdots \\ a(P) \end{bmatrix} = \begin{bmatrix} R^i_{yy}(0) & \cdots & R^i_{yy}(Q_i) \\ \vdots & \ddots & \vdots \\ R^i_{yy}(Q_i) & \cdots & R^i_{yy}(0) \end{bmatrix} \begin{bmatrix} h_i(0) \\ \vdots \\ h_i(Q_i) \end{bmatrix} \tag{8.22}$$

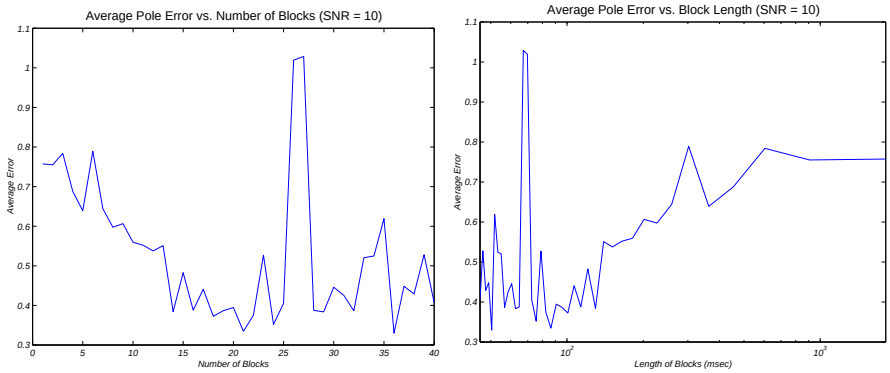
where $i \in \mathscr{M}$ and it has been assumed $Q_i \geq P$. Hence, in each block i, (8.22) can be solved for $\{h_i(q),\ q \in \mathscr{Q}_i\}$.[10] The correlation functions $R^i_{yy}(q)$, $q \in \mathscr{Q}_i$, can be estimated using an equivalent form to the expression in equation (6.8b) of section §6.1.5 (if the amount of data available in each block is small, the *biased* form of the sample autocorrelation function should be used).

[10] It is clear that as $\sigma_w \to 0$, $h_i(q) \to a(p)$, with $\mathscr{Q}_i \equiv \mathscr{P}$: *i.e.* the Wiener-Hopf filter is equivalent to the inverse of the all-pole filter.



(a) Parameter Error vs. Number of Blocks.

(b) Parameter Error vs. Length of Blocks.



(c) Pole Error vs. Number of Blocks.

(d) Pole Error vs. Length of Blocks.

Figure 8.13: *Effect of number of blocks and block length on parameter estimates.*

8.4.3 Effect of Block Length

The effect of the block length is discussed, from a philosophical perspective, in section §8.2.2. In this section, the effect of block length on parameter estimate is considered. Not only does the block length reflect the nonstationarity of the system, as discussed in section §8.2.2, but it also reflects the maximum length of the Wiener-Hopf filter given only a single realisation of the observed data. The results are shown in Figure 8.13, and the figures show that decreasing the block length increases the accuracy of the estimates. Although, for very short block lengths the accuracy decreases (not shown in the figures), there is a wide range of block lengths for which the accuracy is within acceptable limits. Hence, this suggests that only an approximate estimate of the 'optimal' block length is required.

8.5 Temporal Segmentation

The question of choosing the block lengths for the deconvolution model was introduced in section §8.2.2. Currently, in the proposed blind deconvolution model, the block length is heuristically chosen using knowledge of speech characteristics [237]; *i.e.* speech is reasonably well modelled as stationary over periods of around 25 ms.[11] However, as noted in section §8.3, it is not crucial to have detailed knowledge of the source signal model order, as long as the source signal possesses nonstationary statistical properties; even when the source model orders are not known, it is still possible to estimate the channel parameters. Consequently, this raises the question of whether detailed knowledge of block lengths, or the changepoint positions, in the BSAR model is crucial. It is proposed that as long as the statistical properties of the source signal are nonstationary, channel estimation should be possible. The question of choosing block length then reduces to whether its choice yields a stationary or nonstationary source model. As noted in section §8.2.2, if the block length is large, then the variance of the source parameters will be small; however, the model no longer reflects the time-varying nature of the underlying signal and, in the extreme, the source model has stationary statistics. On the other hand, if the block length is small, the source model will have nonstationary characteristics, but the variance of the source estimates will be large, leading to many modes in the posterior distribution for the channel parameters and, hence, to poor parameter estimation.

There exist several Bayesian approaches to segmentation of BSAR processes [95, 283, 298]. These methods assume the observed data is the source signal, rather than a signal filtered by an unknown operator, as in the blind deconvolution problem. However, these methods are easily adapted to account for the channel. Further approaches to the segmentation of nonstationary signals using both parametric and nonparametric models have been considered by Lavielle [208] and Khalil and Duchêne [189]. The approach suggested in section §8.5.2 bares resemblance to the generalised likelihood ratio (GLR) which is widely used in, for example, [20,23,134]; further references are given in [298]. The following methods are provided as motivation for the temporal-spectral changepoint detection problem introduced in section §9.6. Hence, details of the results used in the following

[11] Periods in the range of 20 – 40 ms are common choices for the modelling of speech as stationary.

temporal methods are found in the stated references, and are omitted here since the results in section §9.6 supersede this problem.

8.5.1 Retrospective Changepoint Detection

Retrospective changepoint detection assumes all the data is known, and seeks to estimate the changepoint locations by considering the parameter $\boldsymbol{\tau}$ in equation (7.39a) as unknown random variables. Thus, from (7.39a), where $\boldsymbol{\theta} = \mathcal{I}_{\mathrm{M}}$ is considered as a model dependent function which selects the appropriate number of changepoints, M, and $\mathrm{x} = \mathrm{s}$, it follows (using the notation from section §7.7.3):

$$\mathrm{p}\left(\boldsymbol{\tau}, \mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\Upsilon}, \mathrm{s}, \mathcal{I}\right) = \mathrm{p}\left(\mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\phi}, \mathrm{s}, \mathcal{I}\right) \mathrm{p}\left(\boldsymbol{\tau} \,\middle|\, \mathrm{s}, \mathcal{I}\right) \tag{8.23a}$$

where $\boldsymbol{\Upsilon}$ is the parameter vector $\boldsymbol{\phi}$ with $\boldsymbol{\tau}$ removed; *i.e.* $\boldsymbol{\Upsilon} \triangleq \boldsymbol{\phi}_{-\boldsymbol{\tau}}$. Hence, assigning a non-informative prior to $\mathrm{p}\left(\boldsymbol{\tau} \,\middle|\, \mathrm{s}, \mathcal{I}\right)$ yields:

$$\mathrm{p}\left(\boldsymbol{\tau}, \mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\Upsilon}, \mathrm{s}, \mathcal{I}\right) \propto \mathrm{p}\left(\mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\phi}, \mathrm{s}, \mathcal{I}\right) \tag{8.23b}$$

A MMAP estimate for this expression gives the changepoint locations [283]. The problem of joint changepoint detection and model order selection, with posterior,

$$\mathrm{p}\left(\boldsymbol{\tau}, \boldsymbol{\Xi}, \mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\Gamma}, \mathrm{s}, \mathcal{I}\right) = \mathrm{p}\left(\mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\phi}, \mathrm{s}, \mathcal{I}\right) \mathrm{p}\left(\boldsymbol{\tau} \,\middle|\, \mathrm{s}, \mathcal{I}\right) \mathrm{p}\left(\boldsymbol{\Xi} \,\middle|\, \mathrm{s}, \mathcal{I}\right) \tag{8.23c}$$

where $\boldsymbol{\Gamma} \triangleq \boldsymbol{\phi}_{-\{\boldsymbol{\tau}, \boldsymbol{\Xi}\}}$, can be investigated using Markov chain Monte Carlo (MCMC) methods as discussed in, for example, [298].

8.5.2 Segmentation Decision Ratio

The approach described in this section resembles the widely used generalised likelihood ratio discussed in, for example, [20, 23, 134]. In this method, it is desired to test whether, within a segment of data modelled as a BSAR process, there exists a single changepoint. As such, the probability of a changepoint existing within the segment of data is compared with the probability of there not being a changepoint. To achieve this, suppose the data is denoted by $\{\mathrm{s}(\mathrm{t}),\ \mathrm{t} \in \mathscr{T}'\}$, $\mathscr{T}' = \{\mathrm{t}_0, \ldots, \mathrm{t}_1\}$. A set of hypotheses which describes the different situations are:

Null Hypothesis $\mathcal{I}_0$: The data can be accurately modelled by a stationary AR process across the entire segment of data; *i.e.* a changepoint doesn't exist.

Hypothesis $\mathcal{I}_k$: The data can be accurately modelled by a BSAR process across the data segment, with changepoints at the specified positions $t = t_k$, $t_0 < t_k < t_1$, and with specified AR model order $Q_{k,j}$ in the j-th data block, where $j \in \{1, 2\}$, $k \in \mathscr{K} = \{1, \ldots, K\}$, and K is the number of possible changepoints.

A *superhypothesis* may be defined as a set of all such hypotheses [302]:

$$\mathcal{I} = \{\mathcal{I}_k,\ k \in \mathscr{K}\} \tag{8.24}$$

The problem of strict model selection is the task of inferring the correct model, or hypothesis, and is defined as [302]:

$$\hat{k} = \arg\min_k \mathscr{C}\left(\mathcal{I}_k \,\middle|\, s,\, \mathcal{I}\right) \tag{8.25}$$

where $\mathscr{C}$ is some model-dependent, but parameter independent, criterion. The model for each hypothesis is given by,

$$\begin{aligned} \mathcal{I}_0 :&\quad s(t) = -\sum_{q=1}^{Q} b_0(q)\, s(t-q) + \sigma_0\, \hat{e}(t),\ \forall t \in \mathscr{T}_0 = \{t_0, \ldots, t_1\} \\ \mathcal{I}_{k,\mathbf{Q}_k} :&\quad s(t) = -\sum_{q=1}^{Q_{k,j}} b_{k,j}(q)\, s(t-q) + \sigma_{k,j}\, \hat{e}(t),\ \forall t \in \mathscr{T}_{k,j},\ j \in \{1,2\} \end{aligned} \tag{8.26}$$

where $\mathscr{T}_{k,1} = \{t_0, \ldots, t_k - \tau_k\} \in \mathbb{R}^{T_{k,1}}$, $T_{k,1} = t_k - \tau_k - t_0 + 1$, $\mathscr{T}_{k,2} = \{t_k + 1 + \tau_k, \ldots, t_1\} \in \mathbb{R}^{T_{k,2}}$, $T_{k,2} = t_1 - t_k = \tau_k$, and $T_{k,1} + T_{k,2} = T_0$, $\forall k \in \mathscr{K}$. The parameter τ_k allows for some (unknown) transition band between the stationary models in each block, so as to take account of the case when a process is not truly block stationary. Usually, τ_k is set to some predefined value but, in the general case, a further set of hypothesis can be defined to account for the case when the τ_k's are unknown. Moreover, it is straightforward to implement a further generalisation by defining a sets of hypotheses for the case when the model orders Q and $Q_{k,j}$ are unknown. The criterion function, $\mathscr{C}$, is set to the posterior distribution, since this leads naturally to the MMAP decision criterion discussed later. To derive the posterior distributions, the results in section §7.7.3 are used. The function $f(\mathbf{s}, \boldsymbol{\theta})$, defined in the derivations of equation (7.39), can be considered as hypothesis dependent function, where $\boldsymbol{\theta} = \mathcal{I}_k$, and $\mathbf{x} = \mathbf{s}$, so the Jacobian term is unity. As such, the results presented in the following sections are deftly obtained, where constants of proportionality are retained, and the known parameters are dropped from the expressions for clarity.

8.5.2.1 *Null Hypothesis $\mathcal{I}_0$: No Changepoint in Data Block.*

For the the null hypothesis, $\mathcal{I}_0$, the posterior distribution is given by (7.39b):

$$p\left(\mathcal{I}_0 \,\middle|\, s, \mathcal{I}\right) = \frac{p\left(\mathcal{I}_0 \,\middle|\, \mathcal{I}\right)}{p\left(s \,\middle|\, \mathcal{I}\right)} \frac{\left(\frac{\beta}{2}\right)^{\frac{\nu}{2}} \Gamma(R)}{(2\pi)^{T_0}\, \Gamma\left(\frac{\nu}{2}\right)} \frac{\left\{\gamma + s^T s - s^T S \left(S^T S + \delta^{-2} I_Q\right)^{-1} S^T s\right\}^{-R}}{\left|S^T S + \delta^{-2} I_Q\right|^{\frac{1}{2}}} \tag{8.27}$$

with an appropriate redefinition of the indicies for s and S, and where $R = \frac{\nu+T+1}{2}$.

8.5.2.2 *Hypothesis $\mathcal{I}_k$: Changepoint at Sample t_k*

The posterior distribution for hypothesis $\mathcal{I}_k$ is obtained from equation (7.39a) and is given by:

$$\begin{aligned} p\left(\mathcal{I}_k \,\middle|\, s, \mathcal{I}\right) = \frac{p\left(\mathcal{I}_0 \,\middle|\, \mathcal{I}\right)}{p\left(s \,\middle|\, \mathcal{I}\right)} \prod_{i=1}^{2} \frac{\left(\frac{\beta_{k,i}}{2}\right)^{\frac{\nu_{k,i}}{2}} \Gamma(R_{k,i})}{(2\pi)^{T_{k,i}}\, \Gamma\left(\frac{\nu_{k,i}}{2}\right)} \frac{1}{\left|S_i^T S_i + \delta_{k,i}^{-2} I_{Q_{k,i}}\right|^{\frac{1}{2}}} \\ \times \frac{1}{\left\{\gamma_{k,i} + s_i^T s_i - s_i^T S_i \left(S_i^T S_i + \delta_{k,i}^{-2} I_{Q_{k,i}}\right)^{-1} S_i^T s_i\right\}^{R_{k,i}}} \end{aligned} \tag{8.28}$$

where $k \in \mathscr{K}$ and, again, with an appropriate redefinition of the indicies for the vector s_i and matrix S_i, and $R_{k,i} = \frac{\nu_{k,i}+T_{k,i}+1}{2}$.

8.5.3 Model Selection

The most appropriate model could be found by finding the solution to equation (8.25). However, this can prove computationally expensive, and a method which leads to a computationally efficient sequential algorithm is to use just two hypotheses and the MMAP decision criterion. In this method, a particular changepoint location is chosen heuristically, say hypothesis $\mathcal{I}_1$. The MMAP criterion is then calculated over a window of data, and the window is moved along the entire data sequence one sample at a time. The MMAP ratio is plotted against the position of the hypothesised changepoint within this window, and the MMAP criterion

chooses hypothesis $\mathscr{H}_0$ if:

$$p\left(\mathcal{I}_0 \,\middle|\, s, \mathcal{I}\right) > p\left(\mathcal{I}_1 \,\middle|\, s, \mathcal{I}\right) \tag{8.29a}$$

or, alternatively, if

$$\frac{p\left(\mathcal{I}_0 \,\middle|\, s, \mathcal{I}\right)}{p\left(\mathcal{I}_1 \,\middle|\, s, \mathcal{I}\right)} > 1 \tag{8.29b}$$

In fact, the strict model selection cost function of (8.25) reduces to the MMAP criterion if there are only two hypotheses, $\mathcal{I}_0$ and $\mathcal{I}_1$. The MMAP ratio can be calculated using equations (8.27) and (8.28). However, note that there is no reason to choose different priors for each block and, as such, let $\alpha = \alpha_{1,i}$, $\beta = \beta_{1,i}$ and $\delta = \delta_{1,i}$, $Q = Q_i$ and $T_i = T/2 = \hat{T}$, $\forall i \in \{1, 2\}$. Further, let $\nu = 0, \gamma \to \infty$, such that $\frac{(\frac{\gamma}{2})^{\frac{\gamma}{2}}}{\Gamma(\frac{\gamma}{2})} \to 1$ and, assuming the probability of each model is the same, the left hand side (LHS) of equation (8.29b) reduces to:

$$\frac{p\left(\mathcal{I}_1 \,\middle|\, s, \mathcal{I}\right)}{p\left(\mathcal{I}_0 \,\middle|\, s, \mathcal{I}\right)} = \frac{[\Gamma(R_1)]^2}{\Gamma(R_0)} \frac{\sqrt{|S^T S + \delta^{-2} I_Q|} M^{R_0}}{\sqrt{|S_1^T S_1 + \delta^{-2} I_Q||S_2^T S_2 + \delta^{-2} I_Q|}(M_1 M_2)^{R_1}} \tag{8.30a}$$

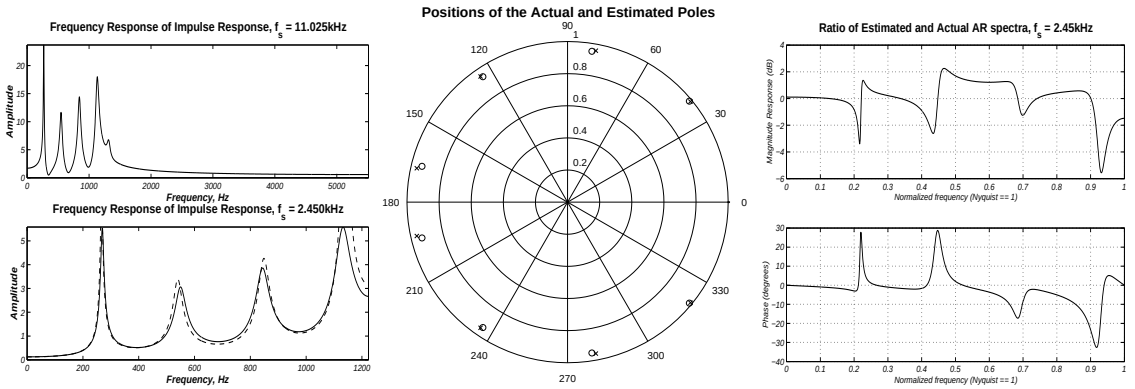
where $R_0 = \frac{2\hat{T}+\nu+1}{2}$ and $R_1 = \frac{\hat{T}+1}{2}$, and

$$\left.\begin{aligned} M &= \gamma + s^T s - s^T S \left(S^T S + \delta^{-2} I_Q\right)^{-1} S^T s \\ M_i &= \gamma + s_i^T s_i - s_i^T S_i \left(S_i^T S_i + \delta^{-2} I_Q\right)^{-1} S_i^T s_i \end{aligned}\right\} \tag{8.30b}$$

As the window is moved along the data sequence, equation (8.30a) can be calculated sequentially by expanding the term $\left(S^T S + \delta^{-2} I_Q\right)^{-1}$ using the *matrix inversion lemma*, or the *Woodbury's formula* [294]. The details of this technique are left as further work.

8.6 Example: Speech Recorded through a Gramophone Horn

As an example of the approach used in this chapter, consider a speech signal which is recorded through a gramophone horn. A typical response [333, 334] of one of these horns is shown in Figure 8.14(a) where the sampling frequency $f_s = 11.025$ kHz. The high-frequency response of this horn is relatively flat, whilst the low-frequency response is resonant and can be accurately modelled by an AR(8)



(a) [Top:] Plot of the true frequency response, $f_s = 11.025$ kHz. [Bottom:] A plot of the true (solid line) and estimated (dotted line) response, $f_s = 2.450$ kHz.

(b) Plot of the true (circle) and estimated (cross) pole positions of the horn IIR model, $f_s = 2.450$ kHz.

(c) Plot of the ratio of the frequency response of the true and estimated response; [Top:] Ratio of magnitude response, [Bottom:] Ratio of phase response.

Figure 8.14: *Estimation of an 8th-order IIR filter using Bayesian method.*

model with $f_s = 2.450$ kHz; this is the frequency response shown in Figure 7.10. The appropriate model orders for the AR processes are assumed to be known, and the block lengths are chosen heuristically. Note that reversible-jump Markov chain Monte Carlo (*rj*-MCMC) algorithms introduced in [129], and used successfully in audio restoration, for example, [363, 364, 378, 379], could be used to tackle the case when the dimensions of the AR models are unknown. This current section is concerned with demonstrating how the information in a nonstationary process can be used to yield solutions to the blind deconvolution problem and, although it is left for future research to investigate the application of *rj*-MCMC techniques to this formulation of blind deconvolution, the investigations and discussion presented in section §8.3 question whether accurate estimation of model orders is *really* necessary.

Taking $\nu_i = \gamma_i = 0, \delta_i$ very large, $Q_i = 50$, $T_i = 300$ and $M = 50$, a distorted speech segment is restored by obtaining a MMAP estimate of $\mathbf{a}$ from (8.7) and deconvolving the observed signal with this estimate. This is achieved by decimating the observed signal to 2.450 kHz and estimating the low frequency response of the horn. The ratio of the true and estimated frequency response of the horn is shown in Figure 8.14(c), where it is seen the estimate is quite accurate. Acoustic listening tests indicate that the restored version is more pleasing to the ear than the speech

heard directly from the horn. The MMAP estimate is obtained by finding the global maximum of $\mathrm{p}\left(\mathbf{a} \,\middle|\, \mathbf{x}, \phi, \mathrm{P}, \mathcal{I}\right)$ which, for a large dimensional pdf is difficult.[12] In this example, the modes were found by searching the parameter space from a variety of starting positions and, in most cases, the largest mode located did not depend on the starting position. The issue of obtaining a global estimate independent of the starting position is discussed in sections §8.2 and §8.3.2 (also see footnote 3 on page 205).

8.7 Chapter Summary

The blind deconvolution problem is tackled by modelling the source signal as a block stationary AR process, and the distortion operator as an IIR filter. A posterior density of the distortion filter parameters conditional on the observed data, the AR model order, and the block lengths, is derived, and an interpretation of the mechanism which allows the blind deconvolution to be performed is discussed. The results of simple examples are given using a MMAP estimate of the filter parameters. The important questions of selecting model order and block length, or changepoint locations, have been discussed: it is noted that by utilising the nonstationarity of the system, less specific knowledge regarding the model of the source signal is required. As long as the model of the source signal is nonstationary, the stationary component of the system can be estimated. If the channel model-order is over-modelled, and the source model order is within a factor of 2 to 3 relative to the true model orders, the channel can be accurately estimated. The effect of observation noise on parameter estimation is considered, and it is shown that use of a block stationary Wiener-Hopf filter to deconvolve the signals improves the accuracy of the parameter estimates.

[12] This optimisation was performed using the Nelder-Mead simplex (direct search) method [294], with an initial condition sufficiently close to the global maximum: *i.e.* as determined by equation (8.9) on page 205.

9

Selective Spectral Modelling

IN Chapter 8, the acoustic system and source signal of Figure 8.1 on page 187 are modelled by a LTI IIR filter and a block stationary AR process, respectively, as discussed in section §7.6. It was tacitly assumed that the all-pole model spectrum spanned the same frequency range as the actual spectrum of the distorting filter, $\mathscr{A}$. However, modelling a real room acoustic impulse response (AIR) by an IIR filter requires a model order in the region of a few hundred to a thousand [262], as discussed in section §6.2. As such, attempting to model the entire acoustic spectrum by a single IIR filter leads to a large computational load, as well as numerical problems due to an enormous parameter space. Moreover, as will be discussed in section §9.3, subband approaches can model nonminimum-phase systems.

The problem of modelling a signal with complicated structure using an all-pole filter is that the model must simultaneously fit the entire frequency range, even though the model may fit some regions in this frequency space better than others. As an example, the spectrum of the acoustic gramophone horn introduced in section §8.6 can be modelled well by an 8th-order AR process in the frequency band $0 - 2.450$ kHz, but requires a *significantly* higher order AR model if the horn spectrum is modelled in the frequency range $0 - 11$ kHz. This suggests it may be better

to model a particular frequency band of the filter's spectrum by an all-pole model, which often results in a lower model order *for that frequency band* and, therefore, improved parameter estimation within this frequency band. In other words, the modelling of different frequency bands has been *decoupled*, and it is proposed that this will lead to a better model fit. Indeed, subband methods have previously been used to model acoustic environments with much success, particularly in the application of dereverberation, as discussed in [9, 206, 261, 375, 384, 385, 403]. Subband linear prediction has been considered by Makhoul [237] and, more recently, by Tan and Fischer [348] and Rao and Pearlman [310]. Additional recent subband approaches to speech enhancement are discussed in [69, 96, 400]. A general discussion of multirate filtering and filter banks may be found in, for example, [373, 374].

9.1 Frequency Domain Formulation

Although the autoregressive model is usually derived in the time-domain, it can also be formulated within the frequency domain. Makhoul [236] suggests such a formulation when analysing speech using linear prediction, and is discussed in more detail in the classic review paper on the subject [237]. Before deriving an approach for selectively modelling a particular frequency band, it is instructive to perform the analysis for the entire spectrum using a frequency domain formulation.

9.1.1 Autocorrelation Method of Least Squares

In the time-domain formulation of the method of least-squares [237, 353], it is sought to minimise the expected value of the square of the excitation sequence for an autoregressive sequence given by:

$$s(t) = -\sum_{p=1}^{P} a(p)\, s(t-p) + e(t), \quad \forall t \in \mathbb{Z} \tag{9.1}$$

where $e(t) \sim \mathcal{N}\left(e(t) \,\middle|\, 0,\, \sigma^2\right)$ and $s(t)$ are the input excitation and resulting output of the process, respectively. These are considered as random processes, although a development assuming deterministic signals leads to a similar result. Furthermore, since the data sequence is assumed to exist over all time, this development is equivalent to the autocorrelation method of parameter estimation. The least-squares

estimate (LSE) is obtained by choosing $\mathbf{a}$, where $[\mathbf{a}]_p = a(p)$, $p \in \{1, \ldots, P\}$, to minimise the quantity $\mathscr{E} = E\left[e(t)^2\right]$. Noting the results from section §3.2.3, the discrete excitation sequence may be written in terms of its Fourier transform such that $e(t) \rightleftharpoons \mathcal{E}(e^{j\omega})$ and, therefore, the mean squared error (MSE) is:

$$\mathscr{E} = E\left[e(t)^2\right] = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E\left[\mathcal{E}(e^{j\omega})\, \mathcal{E}^*(e^{j\omega'})\right] e^{j(\omega-\omega')t}\, d\omega\, d\omega' \tag{9.2a}$$

Taking Fourier transforms of equation (9.1), where $s(t) \rightleftharpoons \mathcal{S}(e^{j\omega})$, gives:

$$\mathcal{E}(e^{j\omega}) = \underbrace{\left[1 + \sum_{p=1}^{P} a(p)\, e^{-j\omega p}\right]}_{A(e^{j\omega})} \mathcal{S}(e^{j\omega}) \equiv A(e^{j\omega})\, \mathcal{S}(e^{j\omega}) \tag{9.2b}$$

where $A(e^{j\omega})$ is the expression indicated above. After substitution into (9.2a):

$$\mathscr{E} = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E\left[\mathcal{S}(e^{j\omega})\mathcal{S}^*(e^{j\omega'})\right] A(e^{j\omega})\, A^*(e^{j\omega'})\, e^{j(\omega-\omega')t}\, d\omega\, d\omega' \tag{9.2c}$$

It can be shown (see Papoulis [286, pp. 418], or section §3.2 of this dissertation), that if $\mathcal{P}_s(e^{j\omega})$ is the power spectrum of a stationary process $s(t)$, then

$$E\left[\mathcal{S}(e^{j\omega})\, \mathcal{S}^*(e^{j\omega'})\right] \equiv 2\pi\, \mathcal{P}_s(e^{j\omega})\, \delta(\omega - \omega') \tag{9.2d}$$

and, as such, equation (9.2c) becomes:

$$\mathscr{E} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{P}_s(e^{j\omega})\, A(e^{j\omega})\, A^*(e^{j\omega})\, d\omega \tag{9.2e}$$

Finally, using the definition of $A(e^{j\omega})$ in (9.2b), after minimising (9.2e) with respect to $a(p)$, it becomes apparent that this is indeed equivalent to the 'standard' normal equations [237, 286, 353], with the autocorrelation terms, $R(p)$, replaced by:

$$R(p) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{P}_s(e^{j\omega})\, \cos(p\omega)\, d\omega \tag{9.3a}$$

This solution also arises if the definition in (9.3a) is substituted into the standard normal equations. When the Fourier spectral components, $\mathcal{S}(e^{j\omega_m})$, are known

only at a finite number of frequencies, $\{\omega_m,\ m \in \{0,\ldots,M-1\}\}$, the power spectral estimate is given by $\mathcal{P}(e^{j\omega_m}) \approx \left|\mathcal{S}(e^{j\omega_m})\right|^2$, and the integral in (9.3a) is replaced by:

$$R(p) = \frac{1}{M}\sum_{m=0}^{M-1} \left|\mathcal{S}(e^{j\omega_m})\right|^2 \cos(p\omega_m) \tag{9.3b}$$

where M is the total number of spectral points on the unit circle, and the frequencies ω_m are those for which a spectral value exists; they need not necessarily be equally spaced. This should be compared with the usual biased expression for the autocorrelation estimate (see [353]) in the time domain,

$$R(p) = \frac{1}{N}\sum_{n=0}^{N-p-1} s(n)\, s(n+p), \quad 0 \le p \le P \tag{9.4}$$

where N is the number of data samples available. Notice that, in both instances, the sequence $s(t)$ is assumed to be zero outside the window $n \in \{0,\ldots,N-1\}$. Therefore, it is clear that if the spectral values, $\mathcal{P}(e^{j\omega_m})$, are already available, there is no computational loss in estimating the parameters, $\mathbf{a}$, using a frequency domain formulation compared to the usual time-domain formulation.

9.1.2 Bayesian Formulation

The frequency domain formulation of the autocorrelation method of least-squares as described in the previous section may seem rather academic, since the end result could have been obtained directly by substituting the time domain estimate for the autocorrelation function, given by equation (9.4), in the normal equations with the spectral estimate in (9.3b). However, the analysis provides insight for developing a spectral parameter estimation technique in the Bayesian framework, which leads naturally onto the technique of selective linear prediction. The frequency domain approach takes a similar line to the time-domain formulation, and the data sequence $\{s(t),\ t \in \mathscr{T} = \{0,\ldots,T-1\}\}$ is assumed to be a windowed version of the infinite sequence introduced in equation (9.1). The discrete Fourier transform (DFT) is defined by [382]:

$$\mathcal{S}(k) \triangleq \mathscr{F}\{s(t)\} = \sum_{t=0}^{T-1} s(t)\, \exp\left\{-\frac{2\pi jkt}{T}\right\}, \quad k \in \mathscr{K} \tag{9.5}$$

where $\mathscr{F}\{\cdot\}$ denotes the DFT and, typically, $\mathscr{K} \equiv \mathscr{T}$.

Application of the DFT to (9.1) for $t \in \mathscr{T}$, with some rearrangement, gives:

$$\mathcal{E}(k) = \mathcal{S}(k) + \sum_{p=1}^{P} a(p)\, \mathcal{S}_p(k) \quad \text{where} \quad \mathcal{S}_p(k) = \exp\left\{-\frac{2\pi\, jkp}{T}\right\} \hat{\mathcal{S}}_p(k) \tag{9.6a}$$

where the modified spectral components $\hat{\mathcal{S}}_p(k)$ are given by:

$$\hat{\mathcal{S}}_p(k) = \sum_{\hat{k} \in \mathscr{K}} \mathcal{S}(\hat{k})\, \mathcal{H}_p(\hat{k} - k) \tag{9.6b}$$

with the precalculated spectral function $\mathcal{H}_p(\hat{k} - k)$ given by:

$$\mathcal{H}_p(\hat{k} - k) \triangleq \frac{1}{T} \exp\left\{\frac{\pi j(\hat{k} - k)(T - p - 1)}{T}\right\} \frac{\sin(\hat{k} - k)\frac{\pi(T-p)}{T}}{\sin(\hat{k} - k)\frac{\pi}{T}} \tag{9.6c}$$

Note that the convolution in equation (9.6b) arises since the process $s(t)$ has been multiplied by a rectangular window such that it is zero outside the range $\mathscr{T}$. An alternative approach, which avoids this convolution, is to assume the signal $s(t)$ is periodic; as T gets larger, the approximation gets better and the results are virtually identical.[1] In this case, equation (9.6c) reduces to the trivial form:

$$\mathcal{H}_p(\hat{k} - k) = \delta(\hat{k} - k) \tag{9.6d}$$

In either case, denoting $\boldsymbol{\mathcal{E}} = \{\mathcal{E}(k),\ k \in \mathscr{K}\}$, equation (9.6a) may be written as:

$$\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{S}} + \mathbf{S}\,\mathbf{a} \tag{9.7}$$

where the matrix $\mathbf{S} = [\boldsymbol{\mathcal{S}}_1\ \cdots\ \boldsymbol{\mathcal{S}}_P]$, with the vector $[\boldsymbol{\mathcal{S}}_p]_k = e^{-\frac{2\pi jkp}{T}} \hat{\mathcal{S}}_p(k), k \in \mathscr{K}$, where $\hat{\mathcal{S}}_p(k)$ is appropriately defined from (9.6b) for each of the spectral responses in (9.6c) and (9.6d). Further, by defining $[\mathbf{W}_T]_{k+1,t+1} \triangleq e^{-\frac{2\pi jkt}{T}}$, $\forall k \in \mathscr{K}$, $\forall t \in \mathscr{T}$, it follows that $\boldsymbol{\mathcal{E}} = \mathbf{W}_T\, \mathbf{e}$. Using the probability transformation [286, 353],

$$p_{\mathcal{E}}(\boldsymbol{\mathcal{E}}) = \frac{1}{|\mathbf{W}_T| \times |\mathbf{W}_T|} p_e\left(\mathbf{W}_T^{-1}\, \boldsymbol{\mathcal{E}}\right) \tag{9.8}$$

and noting $\mathbf{e}$ is WGN, given by $\mathbf{e} \sim \mathcal{N}\left(\mathbf{e} \,\middle|\, \mathbf{0},\ \sigma^2\, \mathrm{I}_T\right)$, and $\mathbf{W}_T$ is square, then:

$$p_{\mathcal{E}}(\boldsymbol{\mathcal{E}}) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}} |\mathbf{W}_T| \times |\mathbf{W}_T|} \exp\left\{-\frac{\boldsymbol{\mathcal{E}}^{\dagger}(\mathbf{W}_T\, \mathbf{W}_T^{\dagger})^{-1}\boldsymbol{\mathcal{E}}}{2\sigma^2}\right\} \tag{9.9}$$

[1] *i.e.* if $P \ll T$, linear convolution can be approximated by circular convolution.

where $\mathbf{z}^\dagger$ denotes the complex-conjugate transpose of the vector or matrix $\mathbf{z}$.[2] Moreover, noting that $\mathbf{W}_T\,\mathbf{W}_T^\dagger \equiv \mathrm{I}_T \;\Rightarrow\; |\mathbf{W}_T| = 1$, where $\mathrm{I}_T \in \mathbb{R}^{T\times T}$ is the identity matrix, equation (9.9) simplifies to:

$$p_{\mathcal{E}}(\boldsymbol{\mathcal{E}}) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \exp\left\{-\frac{\boldsymbol{\mathcal{E}}^\dagger\,\boldsymbol{\mathcal{E}}}{2\sigma^2}\right\} \tag{9.10}$$

as expected since, by Parseval's Theorem in finite-discrete time $\|\mathbf{e}\|^2 = \mathbf{e}^T\,\mathbf{e} = \|\boldsymbol{\mathcal{E}}\|^2 = \boldsymbol{\mathcal{E}}^\dagger\,\boldsymbol{\mathcal{E}}$, where $\|\cdot\|$ denotes the Euclidean norm, and also since the autocorrelation matrix for $\boldsymbol{\mathcal{E}}$ is given by

$$\mathbf{R}_{\mathscr{E}} = \mathbf{W}_T\,\mathbf{R}_e\,\mathbf{W}_T^\dagger = \sigma^2\,\mathbf{W}_T\,\mathbf{W}_T^\dagger = \sigma^2\,\mathrm{I}_T \tag{9.11}$$

Hence, since the Jacobian $\mathscr{J}(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{E}})$ is unity, the likelihood function is:

$$p_{\mathcal{S}}\left(\boldsymbol{\mathcal{S}} \,\middle|\, \mathbf{a}, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \exp\left\{-\frac{(\boldsymbol{\mathcal{S}} + \mathrm{S}\,\mathbf{a})^\dagger\,(\boldsymbol{\mathcal{S}} + \mathrm{S}\,\mathbf{a})}{2\sigma^2}\right\} \tag{9.12}$$

Therefore, using this likelihood with suitable priors for the unknown parameters, a Bayesian analysis can be used to obtain an expression for $p\left(\mathbf{a} \,\middle|\, \boldsymbol{\mathcal{S}}\right)$ and, hence, a parametric estimate for $\mathbf{a}$ given the spectral values $\boldsymbol{\mathcal{S}}$. The maximum-likelihood estimate (MLE) of equation (9.12) is given by: $\hat{\mathbf{a}} = -\left(\mathrm{S}^\dagger\,\mathrm{S}\right)^{-1}\,\mathrm{S}^\dagger\,\boldsymbol{\mathcal{S}}$.

9.2 Selective Subband Modelling

In this section, the frequency domain formulation from the previous section is used for subband modelling. The next subsection discusses the model used for each subband, while sections §9.2.3 and §9.2.5 discuss the spectral-autocorrelation and covariance methods, respectively, for parameter estimation.

9.2.1 Subband Power Spectrum Model

In section §9.1.1 it is shown that the AR parameters for the model of a signal, $s(t)$, can be calculated using the *spectral* components of the signal. This result is used

[2] *i.e.* $\mathbf{z}^\dagger = (\mathbf{z}^*)^T$ where $\mathbf{z}^*$ denotes the complex-conjugate of $\mathbf{z}$.

to inspire the technique of subband modelling. Suppose it is desirable to model the power spectrum, $\mathcal{P}(e^{j\omega}) = \left|\mathcal{S}(e^{j\omega})\right|^2$, of a process $s(t)$, where $s(t) \rightleftharpoons \mathcal{S}(e^{j\omega})$, in the region $\omega_\alpha \leq \omega \leq \omega_\beta$ by an all-pole spectrum. Representing the spectrum as:[3]

$$\mathcal{P}(e^{j\omega}) = \frac{G^2}{\left|1 + \sum_{p=1}^{P} a(p)\, e^{-jp\omega}\right|^2}, \quad \omega \in \Omega_{\alpha\beta} \tag{9.13}$$

where $\Omega_{\alpha\beta} = (\omega_\alpha, \omega_\beta) \subset \mathbb{R}$. However, since the energy is compacted into a small region, high model orders are required. As such, rather than modelling $\mathcal{P}(e^{j\omega})$ at its original subband frequencies, a simpler approach is to model a related signal across the entire spectrum. Consider a signal $\tilde{s}(t)$ whose power spectrum, $\tilde{\mathcal{P}}(e^{j\omega'})$, is related to the spectrum of the original signal, $\mathcal{P}(e^{j\omega})$, by the mapping

$$\tilde{\mathcal{P}}(e^{j\omega'}) \triangleq \mathcal{P}(e^{j\omega}), \quad \omega = \left\{\frac{\omega_\beta - \omega_\alpha}{\pi}\right\}\omega' + \omega_\alpha, \quad \omega' \in (0, \pi) \tag{9.14}$$

Note further that ω' and ω are related by:

$$\omega' = \pi\frac{\omega - \omega_\alpha}{\omega_\beta - \omega_\alpha}, \quad \omega \in \Omega_{\alpha\beta} \tag{9.15}$$

Here, it is seen here that the region $\omega \in \Omega_{\alpha\beta}$ is mapped onto $\omega' \in (0, \pi)$, and the new process, $\tilde{s}(t)$, can be modelled as an all-pole filter across the entire spectrum, with the approximated power spectrum for this signal given by:

$$\tilde{\mathcal{P}}(e^{j\omega}) = \frac{\tilde{G}^2}{\left|1 + \sum_{p=1}^{\tilde{P}} \tilde{a}(p)\, e^{-jp\omega}\right|^2}, \quad \omega \in (0, \pi) \tag{9.16a}$$

Hence, the estimated power spectrum for $\mathcal{P}(e^{j\omega})$, $\omega \in (\omega_\alpha, \omega_\beta)$, is given by:

$$\mathcal{P}(e^{j\omega}) = \frac{\hat{G}^2}{\left|1 + \sum_{p=1}^{\tilde{P}} \tilde{a}(p)\, e^{-jp\pi\frac{\omega-\omega_\alpha}{\omega_\beta-\omega_\alpha}}\right|^2}, \quad \omega \in (\omega_\alpha, \omega_\beta) \tag{9.16b}$$

Observe that this expression is not identical to equation (9.13), and that it generally requires a lower model order than the form in (9.13). Note that the excitation

[3]The assumption that $s(t)$ is an infinite sequence has been made.

variance must be scaled proportionally:

$$\hat{G}^2 = \tilde{G}^2 \frac{\omega_\beta - \omega_\alpha}{\pi} \tag{9.17}$$

since energy must be conserved in this transformation.[4]

The estimated power spectrum for the process $s(t)$ over the complete frequency range, $(0,\ \pi)$, can be represented by a series of subband models, as given by:

$$\left|\mathcal{S}(e^{j\omega})\right|^2 = \sum_{k=0}^{K-1} \frac{\hat{G}_k^2}{\left|1 + \sum_{p=1}^{P_k} a_k(p)\, e^{-jp\pi \frac{\omega-\omega_k}{\omega_{k+1}-\omega_k}}\right|^2} \mathbb{I}_{(\omega_k,\omega_{k+1})}(\omega) \tag{9.19}$$

where the spectrum of the excitation sequence is given by the Fourier transform pair $e(t) \rightleftharpoons \mathcal{E}(e^{j\omega})$, $\mathbb{I}_\Omega(\omega) = 1$ if $\omega \in \Omega$ and zero otherwise, $\omega_0 \triangleq 0,\ \omega_K \triangleq \pi$, and K is the number of subbands.

Although, in the frequency range $\Omega_k = (\omega_k,\ \omega_{k+1})$, equation (9.19) models the power spectrum of the process $s(t)$ and, therefore, the magnitude of the spectrum $\mathcal{S}(e^{j\omega})$, it does not accurately model the phase of $s(t)$, *i.e.* $\arg \mathcal{S}(e^{j\omega})$ since phase information is lost. For a stochastic process, this difficiency is unimportant, since only the power spectrum which needs to be accurately modelled. However, accurately modelling of the phase response is important for channel modelling. This is discussed in the next section.

[4]This result follows by noting the constraint:

$$\int_0^\pi \tilde{\mathcal{P}}(e^{j\tilde{\omega}})\, d\tilde{\omega} = \int_{\omega_\alpha}^{\omega_\beta} \mathcal{P}(e^{j\omega})\, d\omega \tag{9.18a}$$

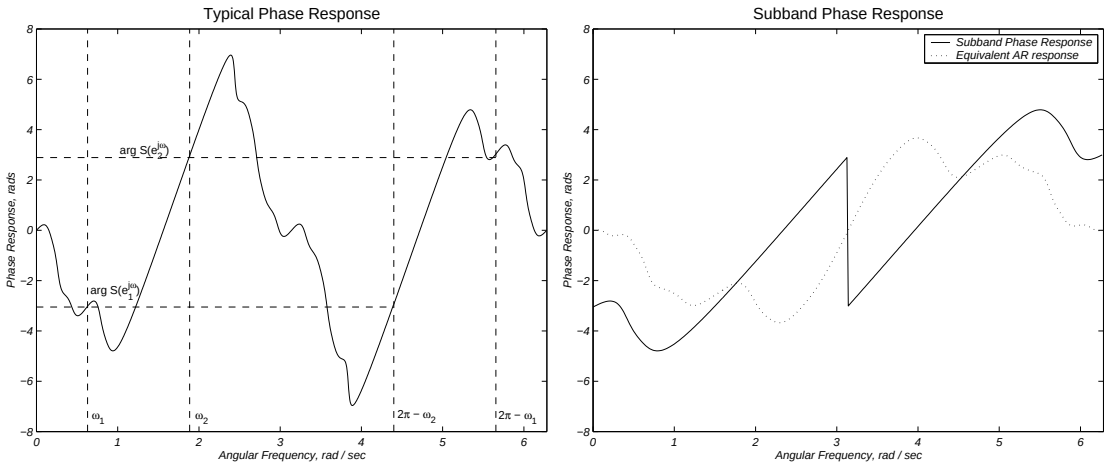
Writing

$$\tilde{\mathcal{P}}(e^{j\tilde{\omega}}) \equiv \frac{\tilde{G}^2}{A(e^{j\omega})} \quad \text{and} \quad \mathcal{P}(e^{j\omega}) \equiv \frac{\hat{G}^2}{A\left(e^{j\pi \frac{\omega-\omega_\alpha}{\omega_\beta-\omega_\alpha}}\right)} \tag{9.18b}$$

then, following the substitution $\hat{\omega} = \pi \frac{\omega-\omega_\alpha}{\omega_\beta-\omega_\alpha}$, it follows:

$$\tilde{G}^2 \int_0^\pi \frac{1}{A(e^{j\omega})}\, d\tilde{\omega} = \frac{\pi \hat{G}^2}{\omega_\beta - \omega_\alpha} \int_0^\pi \frac{1}{A(e^{j\hat{\omega}})}\, d\hat{\omega} \tag{9.18c}$$

and, since the integrals are equivalent, equation (9.17) results.



(a) Typical Phase Response

(b) Subband Phase Response

Figure 9.1: *Demonstrating the Phase Ambiguity.*

9.2.2 Phase Ambiguity

Equation (9.19) suggests that the stochastic process $\{s(t)\}$ is related to its excitation sequence $\{e(t)\}$ by:

$$\mathcal{S}(e^{j\omega}) = \sum_{k=0}^{K-1} \frac{G_k}{1 + \sum_{p=1}^{P_k} a_k(p)\, e^{-jp\pi \frac{\omega-\omega_k}{\omega_{k+1}-\omega_k}}}\, \mathcal{E}(e^{j\omega})\, \mathbb{I}_{(\omega_k,\omega_{k+1})}(\omega) \tag{9.20}$$

where the definitions given with equation (9.19) are used. This model, however, suffers from its inability to model the phase response of the process $s(t)$ and, although this model is sufficient for modelling the power spectrum of a stochastic process, this deficiency causes problems when modelling the channel.

To demonstrate the phase ambiguity, assume in equation (9.20) that, for simplicity, $E(e^{j\omega}) = 1$, $\forall \omega \in \Omega$. Then, it can be seen:

$$\mathcal{S}(e^{j\omega_k}) = \frac{G_k}{1 + \sum_{p=1}^{P_k} a_k(p)} \quad \text{and} \quad S(e^{j\omega_{k+1}}) = \frac{G_k}{1 + \sum_{p=1}^{P_k} (-1)^p\, a_k(p)} \tag{9.21}$$

and, therefore, since $\{a_k(p),\ p \in \mathscr{P}_k\}$ are real coefficients, $\arg \mathcal{S}(e^{j\omega_k}) = 0$ or π. The phase response of a true AR process is always minimum phase. However, consider

the typical phase response of an AR process shown in Figure 9.1(a).[5] The phase of the subband $\Omega_1 \cap \Omega_1' = \{\omega_1,\ \omega_2\} \cap \{2\pi - \omega_2, 2\pi - \omega_1\}$ is shown in Figure 9.1(b), where it is clear that $\arg \mathcal{S}(e^{j\omega_1}) \neq 0$ or π. Hence, the model in equation (9.20) can never model this phase response. A more accurate model must, therefore, be:

$$\mathcal{S}(e^{j\omega}) = \sum_{k=0}^{K-1} \frac{G_k\, e^{j\phi_k(\omega)}}{1 + \sum_{p=1}^{P_k} a_k(p)\, e^{-jp\pi \frac{\omega - \omega_k}{\omega_{k+1} - \omega_k}}} \mathcal{E}(e^{j\omega})\, \mathbb{I}_{(\omega_k, \omega_{k+1})}(\omega) \tag{9.22}$$

where $e^{j\phi_k(\omega)}$ corresponds to an additional phase term to compensate for the difference between the actual phase response, and the phase response of an AR process with identical magnitude responses. Note that the indicator function, $\mathbb{I}_\Omega(\omega)$, represents an *ideal bandpass filter*[6] and, therefore, equation (9.30) is noncausal and can represent nonminimum-phase systems. The additional phase term, $\phi_k(\omega)$, for a particular subband of a minimum-phase systems can often be modelled by a DC offset, a linear term (corresponding to a time-shift) and perhaps a quadratic term. Estimation of this additional phase term is considered in section §9.7.2.

9.2.3 Spectral-Autocorrelation Method

Given the model in equation (9.22), the analysis in section §9.1 can be applied to each subband, $k \in \{0, \ldots, K-1\}$, to obtain estimates of the parameters a_k, provided it is reformulated so that the optimisation is over the frequency range $\omega \in \Omega_k = (\omega_k,\ \omega_{k+1})$. If the Fourier spectral components, $S(e^{j\omega_m})$, of the process $s(t)$ are known only at a finite number of frequencies, $\{\omega_m,\ m \in \{0, \ldots, M-1\}\}$, where the frequencies used in the summation lie strictly in the region being modelled, *i.e.* $\omega_k \leq \omega_m \leq \omega_{k+1}$ where $m \in \{0, \ldots, M-1\}$, then the autocorrelation function of a process, $s_k(t)$, whose power spectrum, $\mathcal{P}_k(e^{j\omega})$, is related to the spectrum of the original signal, $\mathcal{P}(e^{j\omega})$, by the mapping given in equation (9.14) with $\omega_\alpha = \omega_k$ and $\omega_\beta = \omega_{k+1}$, can be estimated using (9.3b). This method is equivalent to the operations outlined in Figures 9.2 and 9.3 and similar to the method in Algorithm 9.1 on the signal $s(t)$. However, this time domain technique does not permit an elegant solution in the Bayesian framework and is computationally more expensive. Note that the operations in Figures 9.2 and 9.3 are simplified for clarity and only show the operations on the positive frequency components. The blocks

[5] This phase response corresponds to the magnitude response shown in Figure 9.7.
[6] See section §4.1 for a discussion of the concept of an ideal filter.

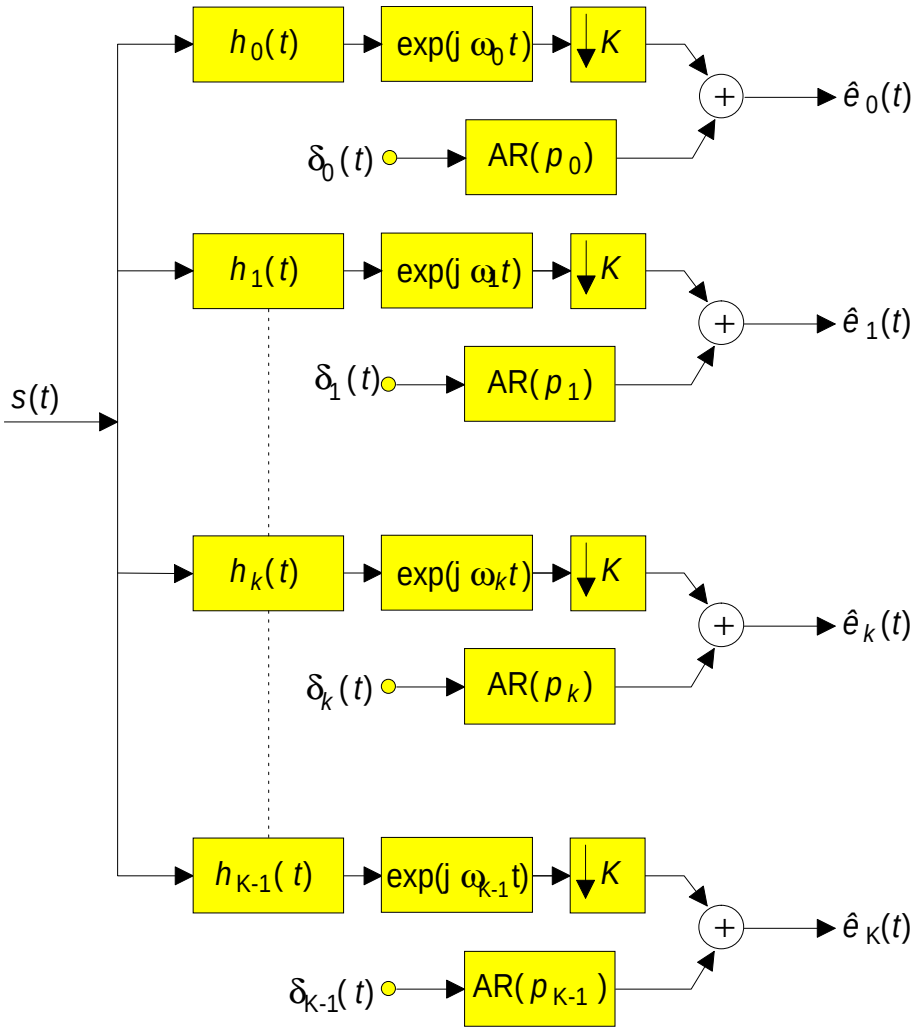


Figure 9.2: *Equivalent subband analysis filter bank (see text).*

indicating AR modelling assume real input and output signals and, thus, also operate on the negative frequency components. The blocks containing $\exp\{\pm j\omega_k t\}$ denote a frequency shift, and the blocks with an '$\downarrow$ K' or '$\uparrow$ K' denote, respectively, decimation and interpolation by a factor of K. Finally, note that $\hat{e}_k$ denotes the AR modelling error and that the excitations are delta functions for channel modelling.[7]

Noting the relationship in (9.7), subband modelling can be performed within a Bayesian framework by simply using the frequency components in the desired

[7]Note that both the temporal and spectral subband AR modelling methods implicitly use a filter bank network and, therefore, care must be taken to ensure that the filter bank possesses perfect reconstruction properties. Details of such techniques are discussed elsewhere [125, 292, 373, 374], and are not taken into account in the current Bayesian analysis for brevity and clarity. However, as the results in section §9.7.1 indicate, perfect reconstruction is a property that must be satisfied.

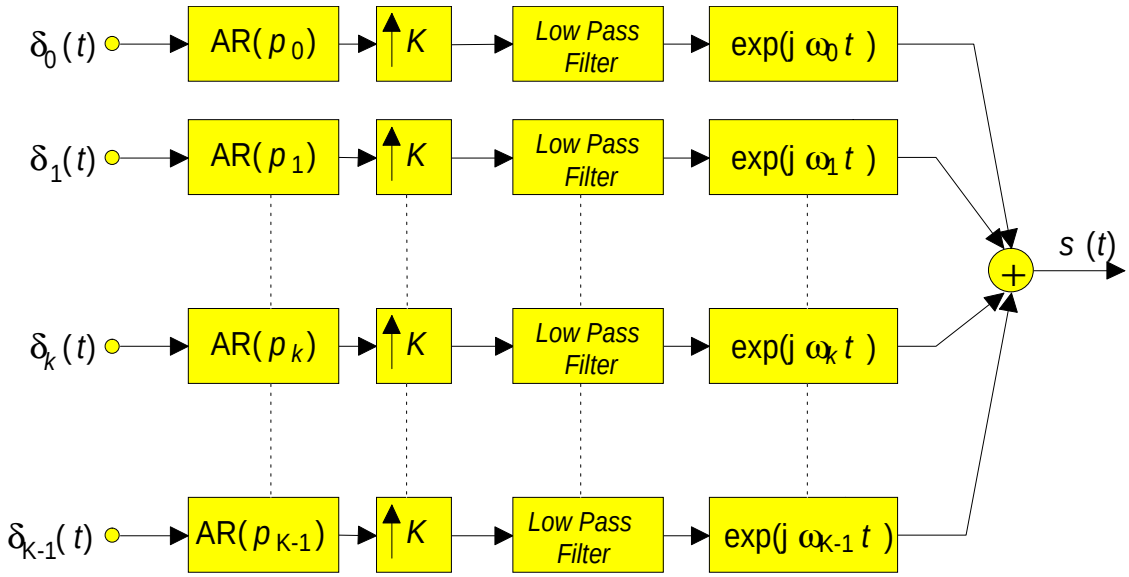


Figure 9.3: *Equivalent subband synthesis filter bank (see text).*

Algorithm 9.1 Equivalence of Subband Modelling.

1: **for** $k = 0$ to $K - 1$ **do**
2: Bandpass filter $s(t)$ in the frequency range $\Omega_k = (\omega_k, \omega_{k+1})$, to give a bandlimited signal $s_k^b(t)$,
3: Modulate the signal $s_k^b(t)$ so that it resides at baseband, to obtain $s_k^m(t)$,
4: Decimate the signal, $s_k^m(t)$, to yield the subsampled signal $s_k^d(t)$,
5: Estimate the AR parameters, $\mathbf{a}_k$, of the signal $s_k^d(t)$,
6: **end for**
7: Reconstruct the spectrum using equation (9.22).

range. The Bayesian analysis in section §9.1.2 can thus be applied on the spectral error sequence $E(e^{j\omega_m})$, $\omega_m \in \Omega_k$, where $\omega_m = \frac{2\pi m}{T}$ and T is the number of error samples corresponding to the sequence $s(t)$. Thus, (9.7) can be directly applied using the appropriate set of spectral components of $s(t)$ corresponding to the region Ω_k in which it is to be modelled.

9.2.4 Zero Extension

If the length of data available is T, and the sequence $s(t)$ is modelled, using the autocorrelation method, by K subbands, then within each subband, there are $\frac{T}{K}$ spectral values available for modelling. If $\frac{T}{K} \leq P$ the problem is degenerate, but even if $\frac{T}{K}$ is relatively small, numerical instability often results, and there may not be

enough data available to estimate the P subband AR parameters. A simple approach for increasing the number of data points in each subband, thus improving numerical stability, when calculating the autocorrelation functions is to zero-pad the data sequence $s(t)$ [382]. To state this formally, consider defining the nonstationary excitation sequence:

$$\hat{e}(t) = \begin{cases} e(t) & \text{for } t \in \mathscr{T} \\ 0 & \text{for } t \notin \mathscr{T} \end{cases} \tag{9.23a}$$

where $e(t)$ is the standard AR excitation WGN sequence. The probability density function for $\hat{e}(t)$ is, consequently, given by:

$$p_{\hat{e}(t)}(\hat{e}(t)) = \begin{cases} p_{e(t)}(\hat{e}(t)) & \text{for } t \in \mathscr{T} \\ \delta(\hat{e}(t)) & \text{for } t \notin \mathscr{T} \end{cases} \tag{9.23b}$$

The excitation sequence over the extended range $\mathscr{T}' = \{0, \ldots, T-1, T, \ldots, T'-1\}$ can, therefore, be written:

$$p_{\hat{\mathbf{e}}}(\hat{\mathbf{e}}) = \prod_{t=0}^{T-1} p_{e(t))}(\hat{e}(t)) \prod_{t=T}^{T'-1} \delta(\hat{e}(t)) \tag{9.24}$$

However, since $\hat{e}(t) = 0$ for $t \in \{T, \ldots, T'-1\}$, the last product of delta's term can be replaced by unity. Using the probability transformation result of equation (9.8), and noting the relationship between the excitation sequence $\hat{\mathbf{e}}$ and its spectral components $\hat{\boldsymbol{\mathcal{S}}}$ given by $\hat{\boldsymbol{\mathcal{E}}} = \mathbf{W}_T \hat{\mathbf{e}}$, then:

$$p_{\hat{\mathcal{E}}}(\hat{\boldsymbol{\mathcal{E}}}) = \frac{1}{|\mathbf{W}_T| \times |\mathbf{W}_T|} p_{\hat{\mathbf{e}}}(\mathbf{W}_T^{-1} \hat{\boldsymbol{\mathcal{E}}}) \tag{9.25}$$

which, by using equation (9.24), the definition of the transformation $\mathbf{W}_T$ matrix, and the pdf for $p_{e(t)}(e(t))$, gives:

$$p_{\hat{\mathcal{E}}}(\hat{\boldsymbol{\mathcal{E}}}) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \exp\left\{ -\frac{\hat{\boldsymbol{\mathcal{E}}}^\dagger \hat{\boldsymbol{\mathcal{E}}}}{2\sigma^2} \right\} \tag{9.26}$$

Moreover, since $e(t)$ and the observed process $s(t)$ are related by equation (9.1), it follows that by extending the excitation sequence, $s(t)$ should also be extended.[8]

[8]There is an edge effect since $s(t)$ depends on previous samples. However, this effect is small if the zero-extension factor is large.

Consequently, by defining $\hat{e}(t)$ as in equation (9.23), a higher number of spectral components can be obtained by creating the sequence,

$$\hat{s}(t) = \begin{cases} s(t) & \text{for } t \in \mathscr{T} \\ 0 & \text{for } t \notin \mathscr{T} \end{cases} \tag{9.27}$$

and substituting the spectral components $\hat{\boldsymbol{\mathcal{S}}} = \mathscr{F}(s)$ into equation (9.12): *i.e.* zero-padding has no influence on the likelihood-function and increases numerical stability. This is as expected since, whilst zero-padding achieves *interpolation*, it does not increase the *spectral resolution* of the discrete Fourier transform [382].

9.2.5 Spectral-Covariance Method

The development of spectral modelling in section §9.1.2 is analogous to the autocorrelation method for AR modelling, and assumes that the data sequence is infinite, or has been 'windowed'. As with the temporal-autocorrelation method, this assumption can lead to *inferior* results, and it would be advantageous to develop a corresponding covariance method. There is, however, a complication which arises in such a formulation: the problem of incorporating initial conditions. In the Bayesian formulation of the spectral-autocorrelation method, the source signal is assumed to be windowed, and this means that any subset of excitation spectral components can be expressed in terms of the AR parameters and source signal spectral components; however, if the sequence is not windowed, the initial conditions must be taken into account. In such a case, there are $(T + P)$ data and T excitation samples and, since there is no longer an isomorphic mapping between the data and the excitation sequence, this does not lend itself to an elegant solution for subband modelling. However, the analysis from sections §9.1.2 and §9.2.3 suggests a method for subband modelling using the temporal-covariance method.

Given the spectral components for the entire data sequence, the components for a particular subband, $\omega_m \in \Omega_{\alpha\beta}$, are obtained by writing:[9]

$$\boldsymbol{\mathcal{S}}' \triangleq \begin{bmatrix} \mathcal{S}(k_\alpha) & \cdots & \mathcal{S}(k_\beta) & \mathcal{S}(T - k_\beta + 1) & \cdots & \mathcal{S}(T - k_\alpha + 1) \end{bmatrix}^T \tag{9.29}$$

[9]The application of subband modelling requires the desired spectral components to be carefully selected. Suppose a subband is to model the spectrum of a T sample sequence in the range $0 \leq \omega_\alpha \leq \omega \leq \omega_\beta \leq \pi$. Taking a T-point DFT yields spectral components at frequencies [382]:

$$\omega_k = \frac{2\pi k}{T}, \quad k \in \mathscr{T} \subset \mathbb{Z} \tag{9.28a}$$

A new (complex-valued) time-domain sequence corresponding to the spectral components in this subband can then be acquired by taking the inverse DFT (IDFT) of the components in $\boldsymbol{\mathcal{S}}'$; $s'(t) \triangleq \mathscr{F}^{-1}(\boldsymbol{\mathcal{S}}')$. The AR parameters for this subband can then be estimated using the covariance method on the sequence $s'(t)$, where the covariance method has been appropriately modified for complex-valued sequences.

9.3 Temporal-Spectral AR Modelling

A natural extension to subband modelling is its application to a data sequence which is block stationary in the time-domain. In this model, the stochastic process $\{s(t)\}$ is related to the excitation sequence $\{e(t)\}$ in block $i \in \mathscr{M}$ by:

$$S_i(e^{j\omega}) = \sum_{k=0}^{K_i-1} \frac{G_{ki}\, e^{j\phi_{ki}(\omega)}}{1 + \sum_{p=1}^{P_{ki}} a_{ki}(p)\, e^{-jp\pi\frac{\omega-\omega_k}{\omega_{k+1}-\omega_k}}} E(e^{j\omega})\, \mathbb{I}_{(\omega_k,\omega_{k+1})}(\omega) \tag{9.30}$$

where the notation used in equation (9.20) of section §9.2.2 is used. In the following sections, the likelihood functions are derived using the spectral-autocorrelation method and then, by drawing analogy with the spectral-autocorrelation method, spectral-covariance formulations are discussed. Note that the Bayesian method discussed so far does not permit the estimation of the additional phase term $\phi_{ki}(\omega)$ in equation (9.30). This section seeks to estimate the subband AR parameters, and the estimation of $\phi_{ki}(\omega)$ is discussed further in section §9.7.2.

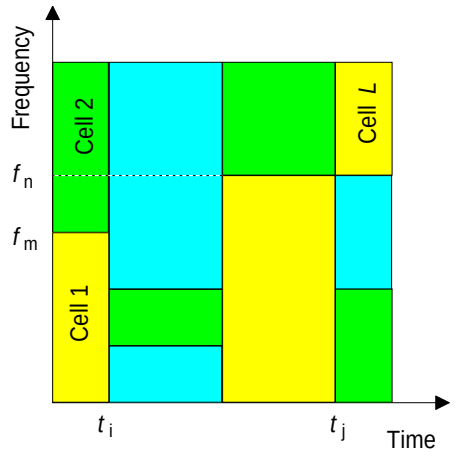
9.3.1 Spectral-Autocorrelation Formulation

In this first formulation, the temporal changepoints are assumed to be at the same point in time for each subband within the corresponding data block: this situation is highlighted in Figure 9.4(a). A similar derivation can be obtained when

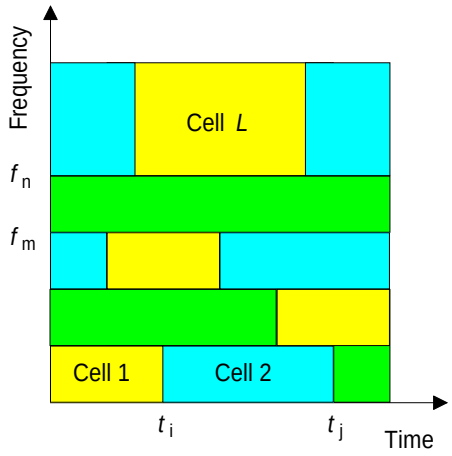
If the signal is bandpass-filtered in the range $\Omega_{\alpha\beta}$ then, since $k \in \mathbb{Z}$, the following components are passed:

$$\begin{aligned} k_\alpha \leq k \leq k_\beta \\ T - k_\beta + 1 \leq k \leq T - k_\alpha \end{aligned} \tag{9.28b}$$

where $k_\alpha = \lceil T\frac{\omega_\alpha}{2\pi} \rceil$, and $k_\beta = \lfloor T\frac{\omega_\beta}{2\pi} \rfloor$. If T is even, it is important to ensure that the component at $\omega_{\frac{T}{2}}$ is not counted twice.



(a) Tiling the time-frequency plane for varying spectral-changepoints over each data block.



(b) Tiling the time-frequency plane for varying temporal-changepoints over each spectral subband.

Figure 9.4: *Systematic tiling of the time-frequency plane.*

the spectral-changepoints remain the same at all time-instances, but the temporal-changepoints depend on the particular subband, as shown in Figure 9.4(b). This may be more useful since different subbands may have different rates of non-stationarity. Suppose the error sequence for a particular model is written as $\mathbf{e} = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_M]$, where there are M data blocks, and $\{\mathbf{e}_i, i \in \mathcal{M}\}$ are the error vectors in each block. Defining the transform:

$$\boldsymbol{\mathcal{E}} = \begin{bmatrix} \boldsymbol{\mathcal{E}}_1 \\ \boldsymbol{\mathcal{E}}_2 \\ \vdots \\ \boldsymbol{\mathcal{E}}_M \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{T_1} & 0 & \cdots & 0 \\ 0 & \mathbf{W}_{T_2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{W}_{T_M} \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{bmatrix} = \mathbf{W}\,\mathbf{e} \tag{9.31}$$

where $\boldsymbol{\mathcal{E}}_i = \mathbf{W}_{T_i}\mathbf{e}_i$, and assuming the excitation sequences are zero-mean white Gaussian noise, $\mathbf{e}_i \sim \mathcal{N}\left(e(t) \,|\, 0, \sigma_i^2\right)$, it follows, after some rearrangement,

$$p_{\mathcal{E}}(\boldsymbol{\mathcal{E}}) = \prod_{i=1}^{M} \frac{1}{\left(2\pi\sigma_i^2\right)^{\frac{T_i}{2}}} \exp\left\{-\frac{\boldsymbol{\mathcal{E}}_i^{\dagger}\boldsymbol{\mathcal{E}}_i}{2\sigma_i^2}\right\} \tag{9.32}$$

Note that the excitation sequence is assumed to have the same variance across each of the subbands: this does not lead to a loss in generality since additional

gain terms can be introduced as in, for example, equation (9.33c). The excitation spectral components in each block, $\boldsymbol{\mathcal{E}}_i$, $i \in \mathscr{M}$, is then be divided into subbands, each of which represent the excitation components for the model in each subband. Hence, writing:

$$\boldsymbol{\mathcal{E}}_i = \left[\boldsymbol{\mathcal{E}}_{0i} \quad \boldsymbol{\mathcal{E}}_{2i} \quad \dots \quad \boldsymbol{\mathcal{E}}_{\{K_i-1\}i}\right] \tag{9.33a}$$

where K_i is the number of subbands in data block i, equation (9.32) becomes

$$p_{\mathcal{E}}(\boldsymbol{\mathcal{E}}) = \prod_{i=1}^{M} \prod_{j=0}^{K_j-1} \frac{1}{\left(2\pi\sigma_i^2\right)^{\frac{T_i}{2}}} \exp\left\{-\frac{\boldsymbol{\mathcal{E}}_{ji}^{\dagger}\boldsymbol{\mathcal{E}}_{ji}}{2\sigma_i^2}\right\} \tag{9.33b}$$

Within each subband, the spectrum is modelled as all-pole and, by using a similar formulation to that in section §9.1.2 and equation (9.7),

$$\boldsymbol{\mathcal{E}}_{ji} = G_{ji}^{-1}\left(\boldsymbol{\mathcal{S}}_{ji} + S_{ji}\,\mathbf{b}_{ji}\right) \tag{9.33c}$$

where G_{ji}^{-1} is the gain of the AR envelope in data block i and subband j. The Jacobian $\mathscr{J}(\boldsymbol{\mathcal{S}}_{ji}, \boldsymbol{\mathcal{E}}_{ji}) = G_{ji}^{T_{ji}}$, where T_{ji} is the number of components in the j-th subband of the i-th data block, and $\sum_{j=0}^{K_i-1} T_{ji} = T_i$. Therefore, if the model parameters are contained in $\boldsymbol{\psi}$, and $\boldsymbol{\vartheta}$ contains model orders, hyperparameters, changepoint locations, and so forth, the likelihood function becomes:

$$p_{\mathcal{S}}\left(\boldsymbol{\mathcal{S}} \,|\, \boldsymbol{\psi}, \boldsymbol{\vartheta}\right) = \prod_{i=1}^{M} \prod_{j=0}^{K_i-1} \frac{1}{\left(2\pi G_{ji}^2\sigma_i^2\right)^{\frac{T_{ji}}{2}}} \exp\left\{-\frac{\left(\boldsymbol{\mathcal{S}}_{ji} + S_{ji}\,\mathbf{b}_{ji}\right)^{\dagger}\left(\boldsymbol{\mathcal{S}}_{ji} + S_{ji}\,\mathbf{b}_{ji}\right)}{2G_{ji}^2\sigma_i^2}\right\}$$

As expected, the gain of the AR envelope and the excitation variance can be combined together to give the usual expression:

$$p\left(\boldsymbol{\mathcal{S}} \,|\, \boldsymbol{\psi}, \boldsymbol{\vartheta}\right) = \prod_{i=1}^{M} \prod_{j=0}^{K_i-1} \frac{1}{\left(2\pi\sigma_{ji}^2\right)^{\frac{T_{ji}}{2}}} \exp\left\{-\frac{\left(\boldsymbol{\mathcal{S}}_{ji} + S_{ji}\,\mathbf{b}_{ji}\right)^{\dagger}\left(\boldsymbol{\mathcal{S}}_{ji} + S_{ji}\,\mathbf{b}_{ji}\right)}{2\sigma_{ji}^2}\right\} \tag{9.34a}$$

The source signal model parameters are now defined as $\boldsymbol{\psi} = \{\mathbf{b}_{ji}, \sigma_{ji}^2, j \in \mathscr{K}_i, i \in \mathscr{M}\}$, where $\mathscr{K}_i = \{0, \dots, K_i - 1\}$ is the number of subbands. If the spectral-changepoints are constant at all time-instances, but the temporal-changepoints de-

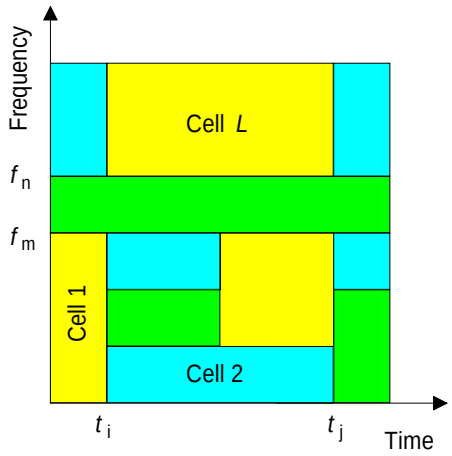


Figure 9.5: *Tiling the time-frequency plane for general temporal-spectral AR modelling.*

pend on the particular subband, the likelihood takes on the more common form:

$$p\left(\boldsymbol{\mathcal{S}} \,\middle|\, \boldsymbol{\theta}\right) = \prod_{j=0}^{K-1} \prod_{i=1}^{M_j} \frac{1}{\left(2\pi\sigma_{ji}^2\right)^{\frac{T_{ji}}{2}}} \exp\left\{ -\frac{\left(\boldsymbol{\mathcal{S}}_{ji} + \mathbf{S}_{ji}\,\mathbf{b}_{ji}\right)^{\dagger}\left(\boldsymbol{\mathcal{S}}_{ji} + \mathbf{S}_{ji}\,\mathbf{b}_{ji}\right)}{2\sigma_{ji}^2} \right\} \tag{9.34b}$$

where the model parameters are $\boldsymbol{\psi} = \{\mathbf{b}_{ji},\, i \in \mathscr{M}_j,\, \sigma_i^2,\, j \in \mathscr{K}\}$, $\mathscr{K} = \{0, \ldots, K-1\}$, and the vectors and matrices containing the spectral components have been suitably redefined: for instance, in subband j, there are M_j temporal-changepoints. Finally, it can be deduced from the form of equations (9.34) that it is possible to tile the time-frequency plane in a general manner, as shown in Figure 9.5. Note that each cell must be rectangular, as it is meaningless to constrain AR parameters to modelling different subbands at different times, which would be the case if the cells were of a non-rectangular shape. The form of the likelihood function can be deduced from (9.34), although this is omitted since the notation becomes rather cumbersome. Further, as discussed in section §9.2.4, improved numerical stability can be obtained by zero-extension of the excitation sequence.

9.3.2 Spectral-Covariance Method

The spectral-covariance method of subband modelling is discussed in section §9.2.5, and it is equivalent to the covariance method applied to a time-domain sequence which corresponds to the spectral components in the required subband. The extension to temporal-spectral AR modelling is as follows: taking the time-

domain sequences for each subband as described in section §9.2.5, find the posterior distribution in a similar form to that derived in section §7.7.3 and, by drawing analogy with the derivation in section §9.3.1, the posterior distribution for the entire data sequence is given by the product of the distributions in each subband, since the subbands are decoupled. The posterior distribution has a similar form to[10] equation (9.34) and, for brevity, is not repeated here. Moreover, the analysis throughout the rest of this chapter assumes that the spectral-autocorrelation method is used, and it is implicitly assumed that the spectral-covariance method is easily obtained, *mutatis mutandis*, from the given results.

9.3.3 Posterior Distribution

The likelihood functions for the two main cases of time-frequency tiling in temporal-spectral modelling are given in equation (9.34). Since each subband within each data block, or *vice versa*, is modelled as an all-pole filter, the prior distributions suggested in sections §7.7.2.1 and §7.7.2.2 are still applicable and, therefore, the following priors are placed on the excitation variance and AR parameters for each time-frequency cell:

$$\begin{aligned} \mathbf{b}_{ji} \,|\, \sigma_{ji}^2,\, \delta_{ji}^2 &\sim \mathcal{N}\left(\mathbf{b}_{ji} \,\middle|\, \mathbf{0}_{Q_{ji}},\, \sigma_{ji}^2\, \delta_{ji}^2\, \mathbf{I}_{Q_{ji}}\right), \quad \delta_{ji} \in \mathbb{R}^+ \\ \sigma_{ji}^2 \,|\, \frac{\nu_{ji}}{2},\, \frac{\gamma_{ji}}{2} &\sim \mathcal{IG}\left(\sigma_{ji}^2 \,\middle|\, \frac{\nu_{ji}}{2},\, \frac{\gamma_{ji}}{2}\right) \end{aligned} \tag{9.35}$$

The development of the posterior distribution from the likelihood function in equation (9.34a) follows the procedure in section §7.7.3 and Appendix D to give:[11]

$$\begin{aligned} p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\mathcal{X}},\, \boldsymbol{\phi},\, \mathcal{I}\right) &\propto p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi},\, \mathcal{I}\right) \\ &\times \prod_{i=1}^{M} \prod_{j=0}^{K_i-1} \frac{\left\{\gamma_{ji} + \boldsymbol{\mathcal{S}}_{ji}^{\dagger}\, \boldsymbol{\mathcal{S}}_{ji} - \boldsymbol{\mathcal{S}}_{ji}^{\dagger} \mathbf{S}_{ji} \left(\mathbf{S}_{ji}^{\dagger}\, \mathbf{S}_{ji} + \delta_{ji}^{-2} \mathbf{I}_{Q_{ji}}\right)^{-1} \mathbf{S}_{ji}^{\dagger}\, \boldsymbol{\mathcal{S}}_{ji}\right\}^{-R_{ji}}}{\left|\mathbf{S}_{ji}^{\dagger}\, \mathbf{S}_{ji} + \delta_{ji}^{-2} \mathbf{I}_{Q_{ji}}\right|^{\frac{1}{2}}} \end{aligned} \tag{9.36}$$

In equation (9.36), $R_{ji} = \frac{T_{ji}+\nu_{ji}+1}{2}$, $j \in \mathcal{K}_i$, $i \in \mathcal{M}$, $\boldsymbol{\phi} \triangleq \{\boldsymbol{\tau},\, \boldsymbol{\omega},\, \boldsymbol{\Xi},\, \boldsymbol{\delta},\, \boldsymbol{\nu},\, \boldsymbol{\gamma}\}$ contains the vector of *changepoints*, or boundaries, of the data blocks, $\boldsymbol{\tau} = \{t_i,\, i \in \mathcal{M}\}$, the vector of subband boundaries, $\boldsymbol{\omega} = \{\omega_{ji},\, j \in \mathcal{K}_i,\, i \in \mathcal{M}\}$, the vector of model orders, $\boldsymbol{\Xi} = \{Q_{ji},\, j \in \mathcal{K}_i,\, i \in \mathcal{M}\}$, and the vectors of hyperparameters,

[10] The differences are that the vectors and matrices contain time-domain components corresponding to the respective subband components, and the Jacobian term $\mathbf{W}_{T_i}$ no longer exists.

[11] Using similar definitions to those in section §7.7.3.

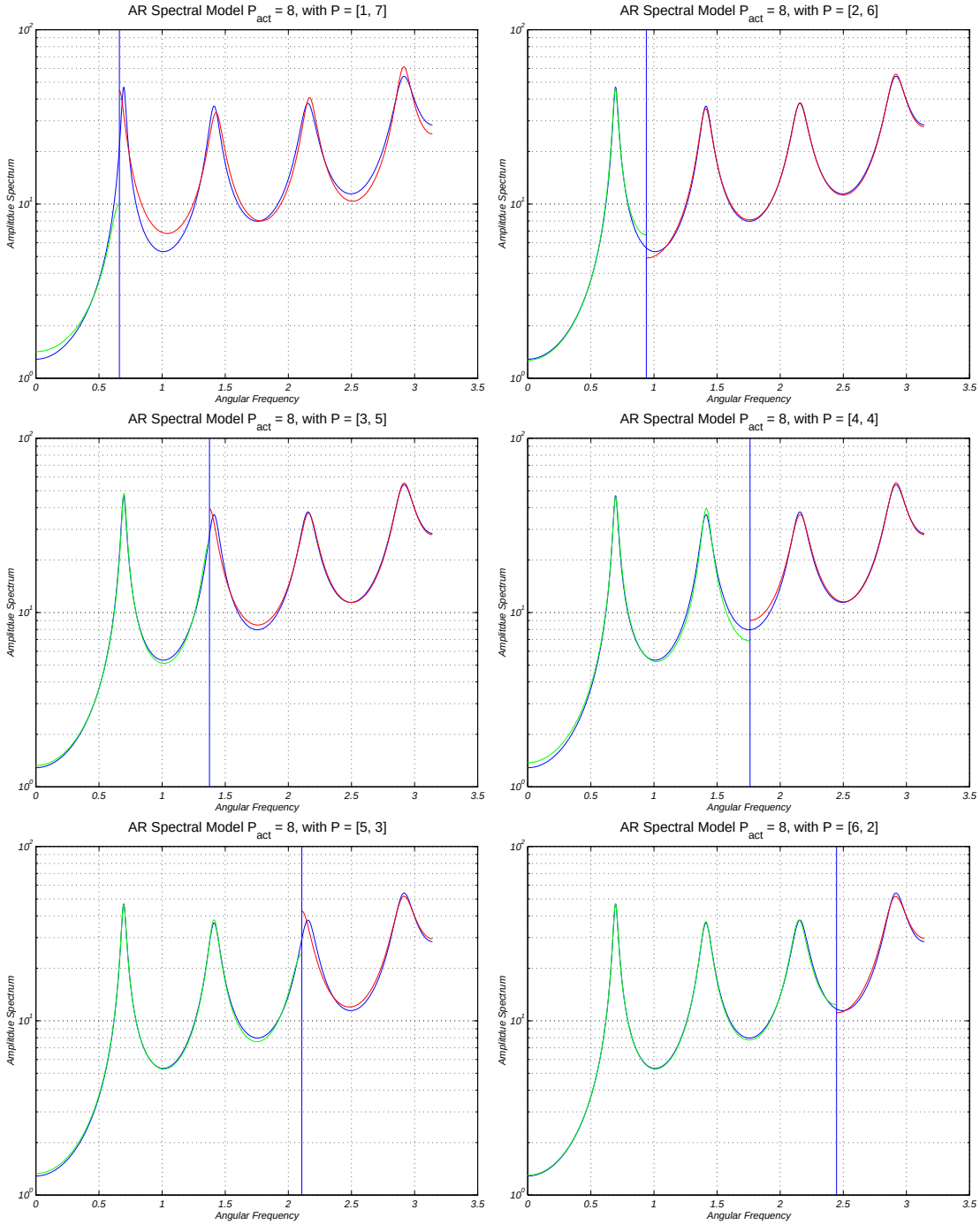


Figure 9.6: *Modelling true AR processes using subband modelling techniques. Each subfigure shows the original spectrum, and the estimated spectrum in each of the two subbands. The vertical line denotes the boundary of the two subbands. This figure shows an AR(8) process modelled using two subbands*

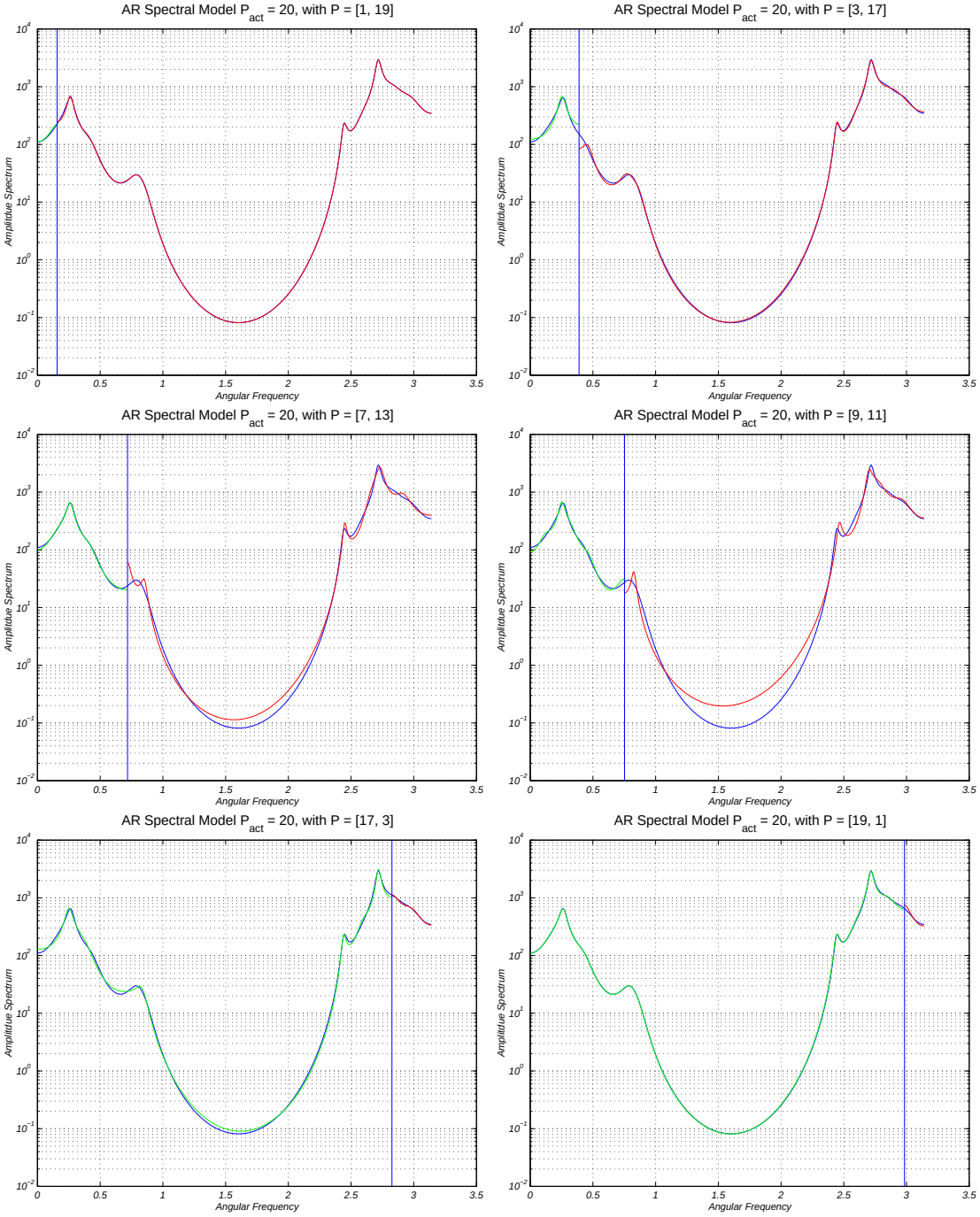


Figure 9.7: *Modelling an AR(20) process using two subbands.*

$\boldsymbol{\delta} = \{\delta_{ji},\, j \in \mathscr{K}_i,\, i \in \mathscr{M}\}$, $\boldsymbol{\nu} = \{\nu_{ji},\, j \in \mathscr{K}_i,\, i \in \mathscr{M}\}$ and $\boldsymbol{\gamma} = \{\gamma_{ji},\, j \in \mathscr{K}_i,\, i \in \mathscr{M}\}$, and $\boldsymbol{\mathcal{X}}$ is a vector of observed spectral components, where $\boldsymbol{\mathcal{X}} = f(\boldsymbol{\mathcal{S}}, \boldsymbol{\theta})$. A similar development follows for the likelihood function in (9.34b). This posterior distribution is used throughout this chapter as (7.39a) is used throughout Chapter 8.

9.4 Subband Modelling Examples

In this section, the subband modelling technique is demonstrated by selectively modelling several *true* AR data sequences in multiple subbands which, together, span the entire frequency range of the spectrum: Rao and Pearlman [310] show examples of modelling a true AR process in subbands. The examples shown in this section demonstrate that there is not necessarily a unique *general* solution to segmenting the spectrum of an AR process. In such an approach, the *model order* for each subband, *changepoint* locations, and the *number of subbands* must be chosen. These can, naturally, be built into a general model and their values estimated by minimising the joint posterior distribution for all these variables using *Markov chain Monte Carlo* (MCMC) methods [129]. However, as with all such approaches, this may yield many *local* solutions, and it is important to have an understanding of the type of solutions that may result.

In the first example, the gramophone horn introduced in section §8.6 is modelled using two subbands. The model order in each band is fixed, and chosen such that the sum across both bands equals the model order of the horn across the entire frequency range. Since the number of subbands and the model order in each subband is constrained, the location of the spectral changepoint can be determined using, for example, a MMAP estimate, and the details of how this is achieved is found in section §9.6. Figure 9.6 shows the estimated spectra in each subband for various choices of model order. Figure 9.6 also shows similar results for a 20-th order AR filter. Notice that for the various model orders, the changepoint is never placed near the center of the trough of the spectrum. This is since AR models generally are better at modelling resonants than nulls and, therefore, a low order process cannot model a null near a changepoint.

This therefore suggests that when a particular spectrum is divided into subbands, the *model orders* for each subband should be selected *a priori* by some means, and the subband boundaries located using the method discussed in sec-

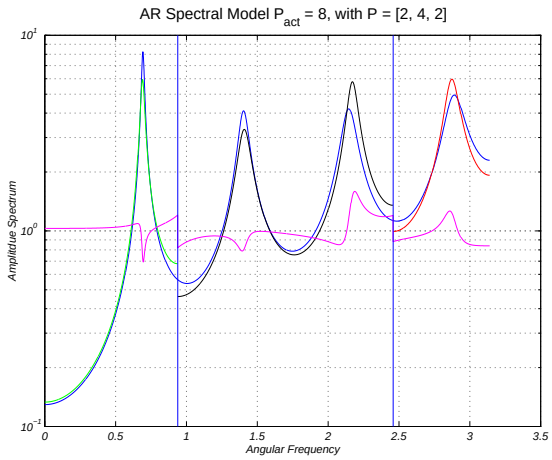


Figure 9.8: *Blind deconvolution using subband modelling.*

tion §9.6, rather than choosing the subband locations and estimating the model order or, equally, rather than choosing *both* the model order and subband widths.

9.5 Subband Blind Deconvolution

The blind deconvolution problem in section §8.2 is now revisited by considering a subband approach. Suppose that the distortion filter, $\mathscr{A}$, is modelled in subbands using all-pole filters. The input of the system, $s(t)$, is related to the output, $x(t)$, of the system by $x(t) = f(s(t), \boldsymbol{\theta})$ or, equivalently, $X(e^{j\omega}) = A(e^{j\omega})\, S(e^{j\omega})$ where:

$$A(e^{j\omega}) = \sum_{i=0}^{K_i-1} \frac{1}{a_i(0) + \sum_{p=1}^{P_i} a_i(p)\, e^{-jp\pi\frac{\omega-\omega_i}{\omega_{i+1}-\omega_i}}} \mathbb{I}_{(\omega_i,\omega_{i+1})}(\omega)\ , \tag{9.37}$$

and where $\omega_0 \triangleq 0$, $\omega_K \triangleq \pi$, and K is the number of subbands. Note that the $a_i(0)$ term defines the gain of the filter and, in order to avoid scaling ambiguity, $a_1(0) \triangleq 1$. If the boundaries of the frequency bands ω_i, $i \in \mathscr{K} = \{0, \ldots, K\}$, are aligned with the subband boundaries of the source model in equation (9.30) then, extending the method discussed in section §8.2, the posterior distribution for the AR parameters in each subband, $\mathscr{A} \triangleq \{\mathbf{a}_i, i \in \mathscr{K}\}$, where $\mathbf{a}_i = \{a_i(p), p \in \mathscr{P}_i = \{0, \ldots, P_i\}\}$ for $i \in \{2, \ldots, K\}$ and $\mathbf{a}_1 = \{a_1(p), p \in \mathscr{P}_1 = \{1, \ldots, P_1\}\}$, is the analogue of

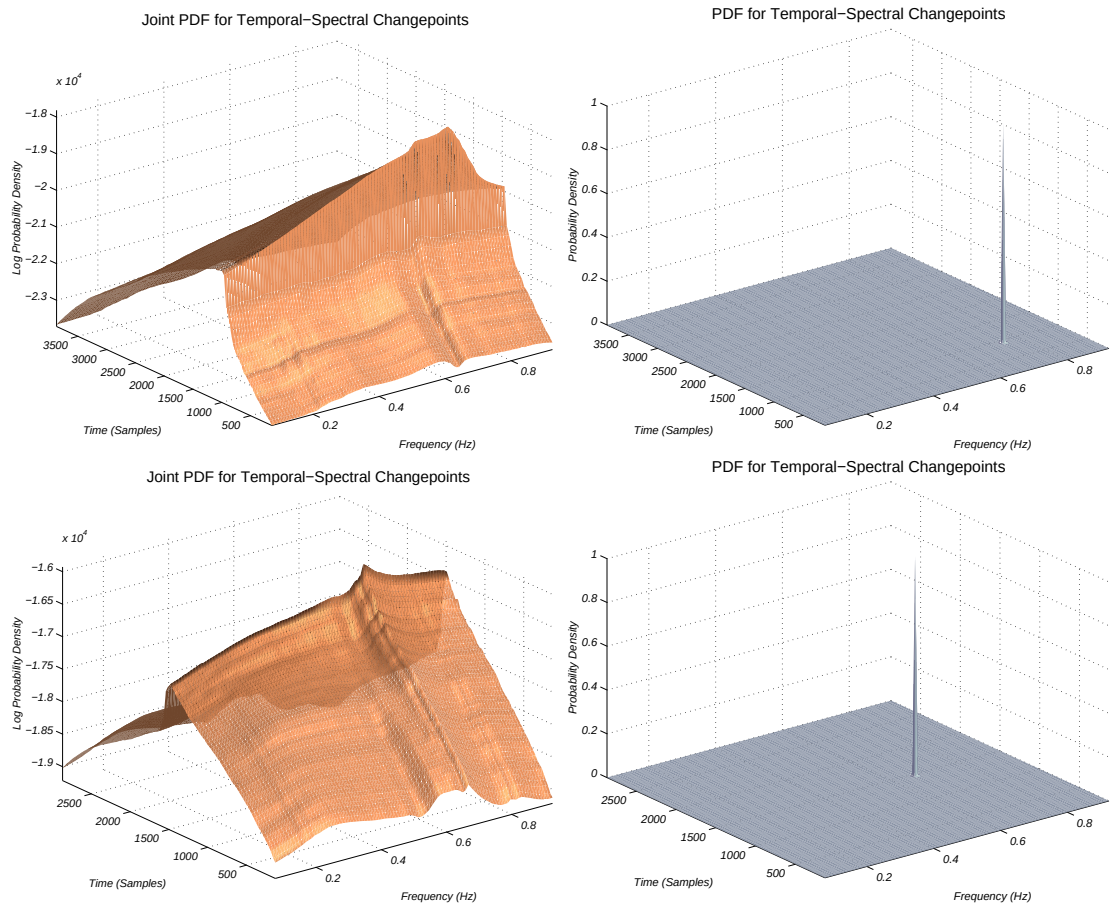


Figure 9.9: *The posterior densities for estimating the temporal-spectral changepoint of a 2nd order TVAR process, where there are two data blocks and two subbands. The two rows are for different block lengths, and different bandwidths.*

equation (8.7) and, with all the usual definitions, is given by:

$$\mathrm{p}\left(\mathscr{A}, \mathbf{P} \,\middle|\, \boldsymbol{\mathcal{X}}, \boldsymbol{\phi}, \mathcal{I}\right) \propto \frac{\mathrm{p}\left(\mathscr{A}, \mathbf{P} \,\middle|\, \boldsymbol{\phi}, \mathcal{I}\right)}{\mathscr{J}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{S}}\right)} \times \prod_{i=1}^{M} \prod_{j=0}^{K_i-1} \frac{\left\{\gamma_{ji} + \boldsymbol{\mathcal{S}}_{ji}^{\dagger}\boldsymbol{\mathcal{S}}_{ji} - \boldsymbol{\mathcal{S}}_{ji}^{\dagger}\mathbf{S}_{ji}\left(\mathbf{S}_{ji}^{\dagger}\mathbf{S}_{ji} + \delta_{ji}^{-2}\mathbf{I}_{Q_{ji}}\right)^{-1}\mathbf{S}_{ji}^{\dagger}\boldsymbol{\mathcal{S}}_{ji}\right\}^{-R_{ji}}}{\left|\mathbf{S}_{ji}^{\dagger}\mathbf{S}_{ji} + \delta_{ji}^{-2}\mathbf{I}_{Q_{ji}}\right|^{\frac{1}{2}}} \tag{9.38}$$

where $R_{ji} = \frac{T_{ji}+\nu_{ji}+1}{2}$, $j \in \mathscr{K}_i$, $i \in \mathscr{M}$, and $\mathbf{P} = \{P_i, i \in \mathscr{K}\}$. As an example, the 8th-order gramophone horn model shown in Figure 7.10(b) is driven by a BSAR(12) source signal model as discussed in section §7.6.1.2. The estimated channel, found by maximising equation (9.38) using the Nelder-Mead simplex

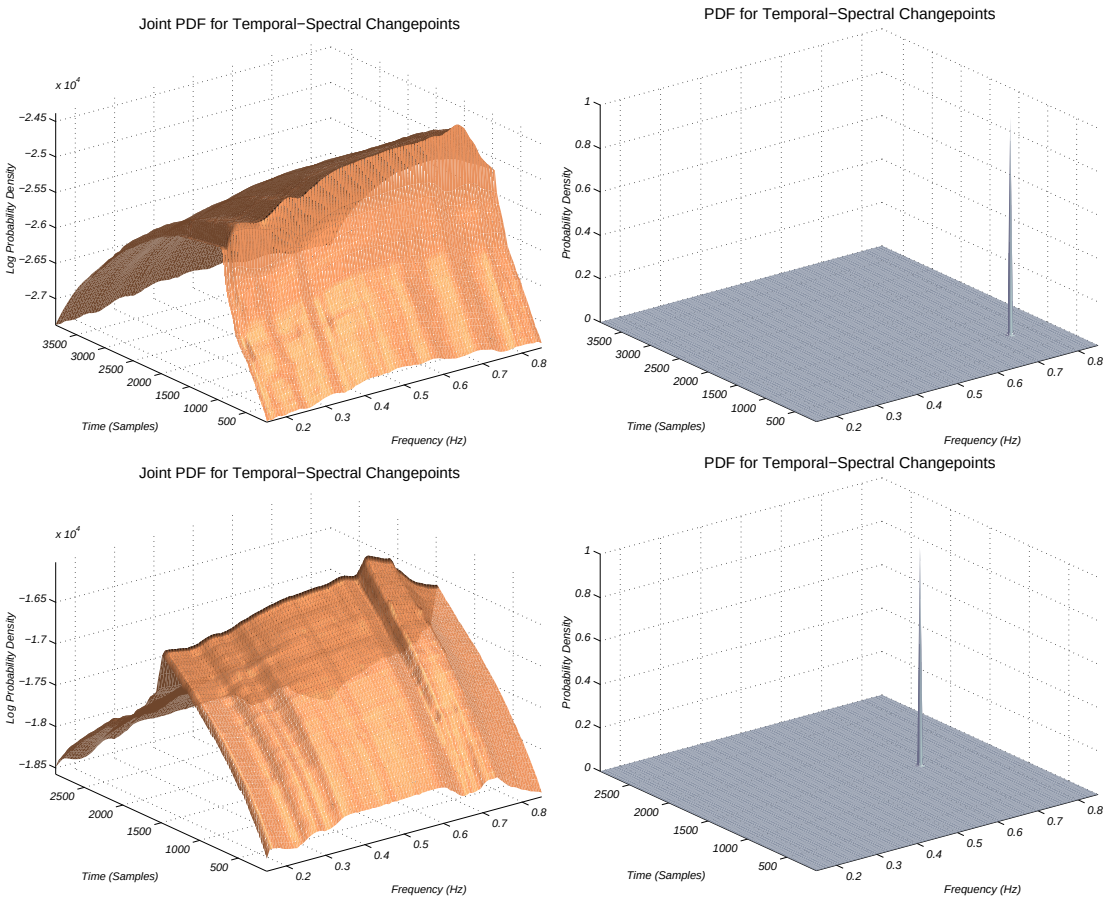
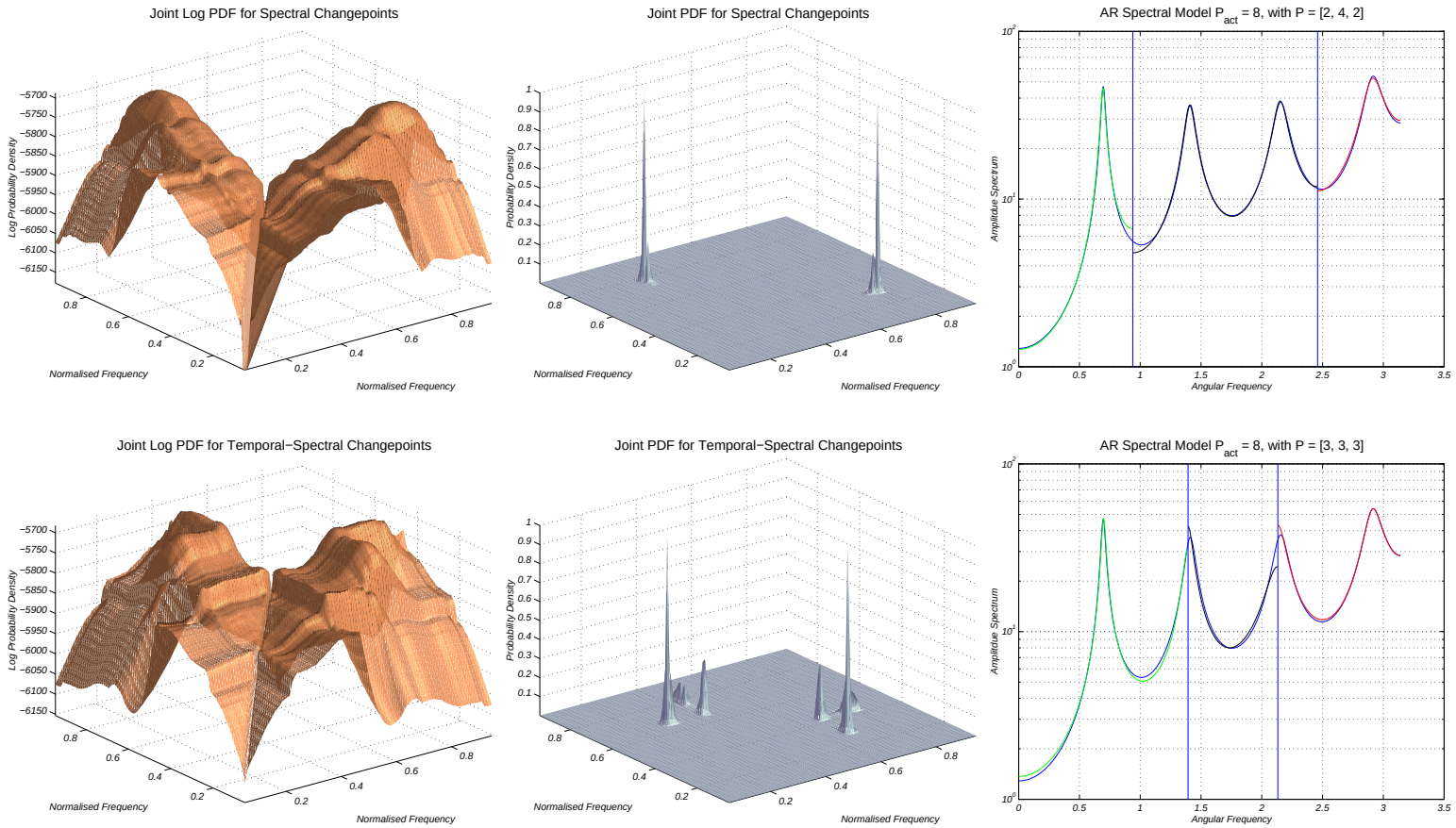


Figure 9.10: *The posterior densities for estimating the temporal-spectral changepoint of an 8th-order TVAR process, where there are two data blocks and two subbands. The two rows are for different block lengths, and different bandwidths.*

method [294] (see footnote 3 on page 205), is shown in Figure 9.8. Apart from slight discontinuities at the subband boundaries, the resulting estimate is accurate.

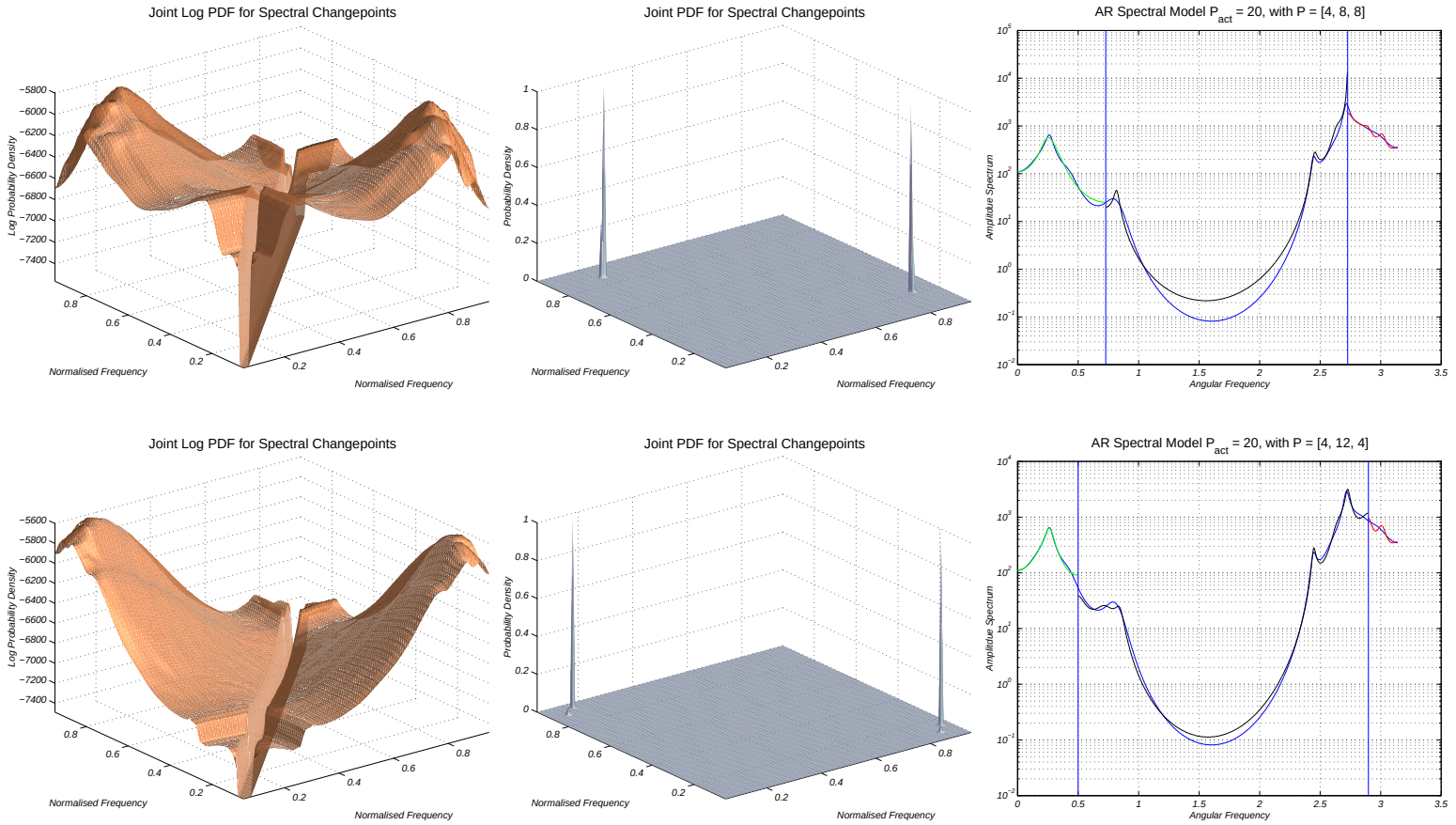
9.6 Retrospective Temporal-Spectral Changepoint Detection

In section §8.5, the problem of locating the boundaries of the homogeneous data segments was discussed. This section considers the problem of retrospective



(a) Modelling an AR(8) process using two subbands.

Figure 9.11: Left Column: *The log-pdf for the positions of the two subband boundaries, and* Center Column: *the pdf, emphasising the sharpness of the peaks.* Right Column: *The corresponding estimated spectrum using least-squares estimate.*



(a) Modelling an AR(20) process using two subbands.

temporal-spectral changepoint detection[12] where it is assumed that all the data is available for processing. Since the data block and subband boundaries will not be known in a completely blind system, it is of great interest to determine whether these changepoints can be detected. If there is a single data block, and a number of subbands, their boundaries can be determined using a MMAP decision criterion similar to the one developed in section §8.5.2. However, this methodology is not considered here since, in general, there will be more than one data-block and one subband. Therefore, a method similar to that developed in section §8.5.1 is investigated. Considering the parameters $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$ in equation (9.36) as random variables, where $\boldsymbol{\theta} = \mathcal{I}_{\mathrm{M}}$ is considered as a model dependent function which selects the appropriate number of blocks and subbands, and $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}}$, then

$$\mathrm{p}\left(\boldsymbol{\tau},\, \boldsymbol{\omega},\, \mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\Upsilon},\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right) = \mathrm{p}\left(\mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\phi},\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right)\, \mathrm{p}\left(\boldsymbol{\tau} \,\middle|\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right)\, \mathrm{p}\left(\boldsymbol{\omega} \,\middle|\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right) \tag{9.39}$$

where $\boldsymbol{\Upsilon} \triangleq \boldsymbol{\phi}_{-[\boldsymbol{\tau},\boldsymbol{\omega}]}$. Hence, by assigning a non-informative prior to $\mathrm{p}\left(\boldsymbol{\tau} \,\middle|\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right)$ and $\mathrm{p}\left(\boldsymbol{\omega} \,\middle|\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right)$, the posterior distribution for the changepoint locations is found as:

$$\begin{aligned}
&\mathrm{p}\left(\boldsymbol{\tau},\, \boldsymbol{\omega},\, \mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\Upsilon},\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right) \propto \mathrm{p}\left(\mathcal{I}_{\mathrm{M}} \,\middle|\, \boldsymbol{\phi},\, \boldsymbol{\mathcal{S}},\, \mathcal{I}\right) \\
&\quad \propto \prod_{i=1}^{M} \prod_{j=0}^{K_i-1} \frac{\left\{\gamma_{ji} + \boldsymbol{\mathcal{S}}_{ji}^{\dagger}\boldsymbol{\mathcal{S}}_{ji} - \boldsymbol{\mathcal{S}}_{ji}^{\dagger}\mathbf{S}_{ji}\left(\mathbf{S}_{ji}^{\dagger}\mathbf{S}_{ji} + \delta_{ji}^{-2}\mathbf{I}_{Q_{ji}}\right)^{-1}\mathbf{S}_{ji}^{\dagger}\boldsymbol{\mathcal{S}}_{ji}\right\}^{-R_{ji}}}{\left|\mathbf{S}_{ji}^{\dagger}\mathbf{S}_{ji} + \delta_{ji}^{-2}\mathbf{I}_{Q_{ji}}\right|^{\frac{1}{2}}}
\end{aligned} \tag{9.40}$$

The following section discusses some examples of temporal-spectral segmentation.

9.6.1 Segmentation Examples

Figures 9.9 through 9.11(a) show examples of spectral-temporal segmentation. The results, as a whole, demonstrate that subband modelling is a very promising technique, and should lead to improved results for the blind deconvolution problem. In summary:

Figure 9.9 and Figure 9.10 show spectral-temporal segmentation of process which is block stationary, and is truly a subband AR process: in other words, each part of the spectra is modelled by equation (9.30). The temporal and spectral changepoints for the process are estimated using the method described in section §9.6, and the results are very close to the true temporal and spectral

[12] Alternatively referred to as temporal-spectral segmentation.

changepoints. This demonstrates from a theoretical viewpoint the validity of the subband model.

Figure 9.12(a) attempts to model a standard AR(20) model using the subband technique. The underlying process is truly autoregressive, and it is of interest to see how the method splits the spectra into the three subbands; further details of the performance of this modelling is discussed in section §9.4. What is of interest here, is the *sharpness* and *confidence* in the log-pdf and pdf of the changepoint locations. This indicates that, given a set of subband model orders, a spectra can be divided *confidently* into subbands. A similar result is given in Figure 9.11(a) for an AR(8) process.

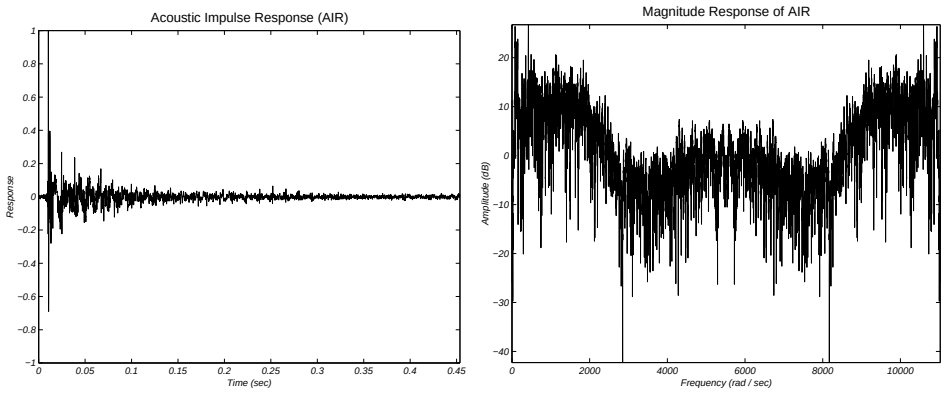
9.7 Subband Modelling of Room Acoustics

If the subband autoregressive model is an appropriate representation for a particular system, then it elegantly facilitates blind system identification. The primary application in this dissertation is the cancellation of reverberation in acoustic environments. Thus, this section investigates the suitability of the subband model to acoustic dereverberation. As discussed in section §6.3.1, many authors have attempted subband approaches to dereverberation with mixed success. The subband model introduced in equation (9.22) in section §9.2.1 is used to represent a *known* acoustic impulse response. Naturally, in practice the AIR will be unknown: however, the purpose of this section is to investigate the ability of the subband model to equalise the room transfer function. Blind deconvolution of speech in acoustic environments is considered in section §9.8. Due to the problems associated with estimating the phase response of a system, the equalisation of the magnitude and phase responses are considered separately in sections §9.7.1 and §9.7.2 respectively.

9.7.1 Reconstructing the Magnitude Frequency Response

A typical impulse response, measured in a stairwell, with no direct path from source to observer, is shown in Figure 9.12(b), with magnitude frequency response shown in Figure 9.12(c).[13] The length of this impulse response is $T = 4000$ samples. The

[13] This response is the same as that shown in Figure 6.3(b) on page 129.



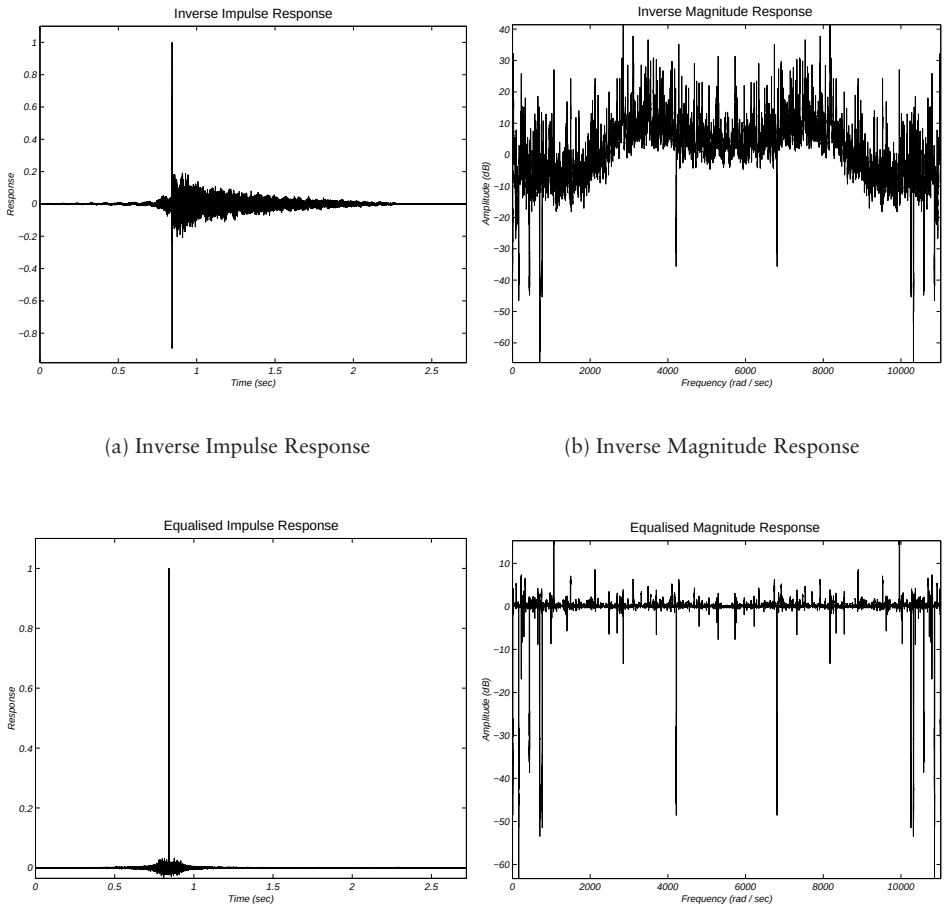
(b) Impulse Response (c) Magnitude Frequency Response

Figure 9.12: *Typical acoustic impulse response: source and observer in a stairwell with no direct path.*

minimum-phase equivalent of this impulse response, calculated using the method discussed in section §6.3.3.3, is modelled using the subband representation in equation (9.22) using a ML estimate calculated from the spectral-AR method discussed in section §9.2.3, where the phase response is modified as discussed in the next section. The model order in each subband is $P = 50$, there are $K = 500$ subbands, and the zero-extension factor, Z, is such that the length of the extended impulse response is $T' = TZ \equiv 2^N$ for $N \in \mathbb{Z}^+$. Modelling of the nonminimum-phase response is briefly discussed at the end of this section. Although the minimum-phase impulse response can be inverted directly, in blind deconvolution applications it is infeasible to estimate an AIR of length $T = 4000$ using a single AR model. Hence, to investigate the suitability of the subband model, which can be easily estimated using a number of low-order optimisations, the inverse impulse response is calculated by inverting the frequency response in each subband and taking the inverse Fourier transform. The results are shown in Figure 9.13(a), with the magnitude response in Figure 9.13(b) (compare with Figure 9.12(c)).

Figure 9.13(c) shows the convolution[14] of the AIR in Figure 9.12(b) and the inverse response in Figure 9.13(a). The magnitude frequency response of the equalised room transfer function shown in Figure 9.13(d) indicates that the spectral coloration is significantly reduced. However, as demonstrated in Figure 9.14, a closer inspection reveals why the magnitude response contains many sharp spectral

[14]This convolution is performed in the frequency domain to reduce computation and improve numerical accuracy.



(a) Inverse Impulse Response

(b) Inverse Magnitude Response

(c) Equalised Impulse Response

(d) Equalised Magnitude Response

Figure 9.13: *Inverse response to the room impulse response shown in Figure 9.12.*

components: since the model in each subband is completely decoupled from the other subbands, there are discontinuities in the spectrum at the subband boundaries. The model in equation (9.22) does not enforce any continuity between blocks. Since the original spectrum is continuous, this is a short-coming of the model. An approach to ensuring continuity between subband boundaries is to modify the prior distributions for the AR parameters such that the end point at the lower subband boundary is constrained to match the estimated spectrum in the previous subband: such a technique is left for future work. It is also noted that the resonant spikes at the boundaries of the subbands in Figure 9.14 are due to the fact that the model parameters have been estimated using a filter bank which does not possess perfect reconstruction properties. This issue is left as further work, and the

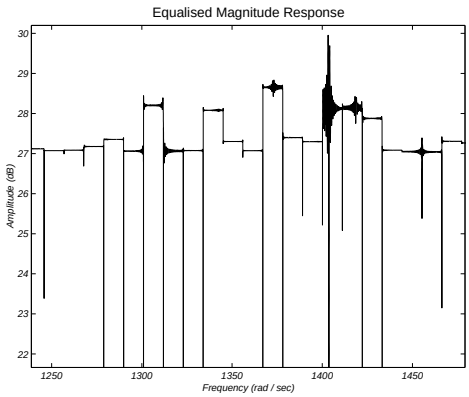


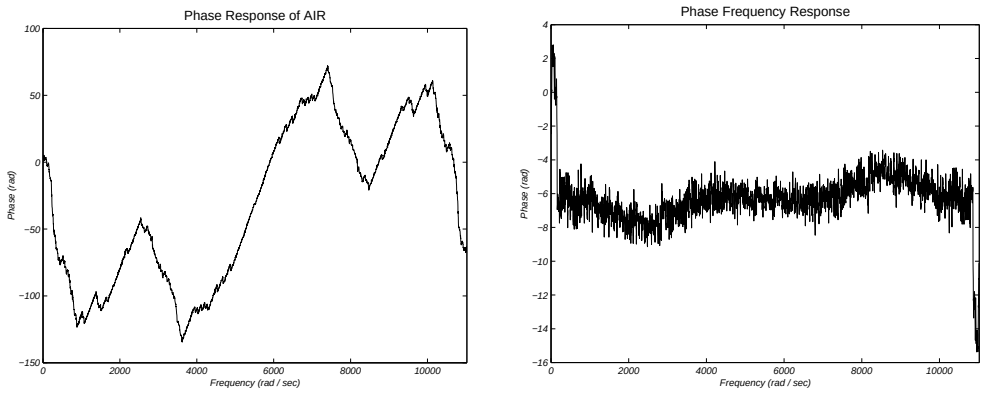
Figure 9.14: *A close inspection of the equalised magnitude response indicates that there are discontinuities between the subbands.*

details of such filter banks may be found in, for example, [125, 292, 373, 374].

The modelling of the magnitude frequency response of the nonminimum-phase impulse response gives similar results to the minimum-phase system, since the magnitude responses are identical, although there are slightly larger discontinuities. The problem with nonminimum-phase system is the modelling of the phase response as discussed in the next section, and the subsequent inversion: *i.e.* a noncausal inverse. This issue is discussed further in section §9.8.

9.7.2 Reconstructing the Phase Frequency Response

Since least-squares or Bayesian autoregressive parameter estimators minimise the MSE of the spectral error function (see section §9.1), they cannot model the additional phase term $\phi_k(\omega)$ in equation (9.22) and, therefore, modelling the phase response of an acoustic impulse response is made more difficult. The (nonminimum) phase response corresponding to the AIR in Figure 9.12 is shown in Figure 9.15(a), and the minimum-phase equivalent response is shown in Figure 9.15(b). In modelling the impulse response shown in Figure 9.12(b), it was sought to minimise the phase discrepancy introduced when using a parameter estimate based on the spectral error function. This is achieved by assuming that linear and quadratic phase terms contribute significantly to the additional phase. Modelling the additional phase terms in such a way accounts for the phase-shift obtained by virtue of the subband approach, as discussed in section §9.2.2, without conflicting with the phase contribution of the autoregressive estimator. Hence, an estimate of $\phi_k(\omega)$ is



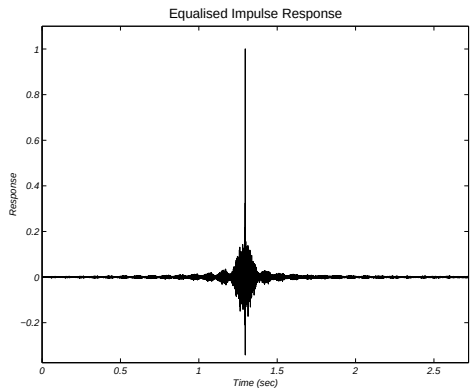
(a) Phase Frequency Response (b) Phase of Minimum-Phase Equivalent Response

Figure 9.15: *Phase responses of acoustic impulse response shown in Figure 9.12.*

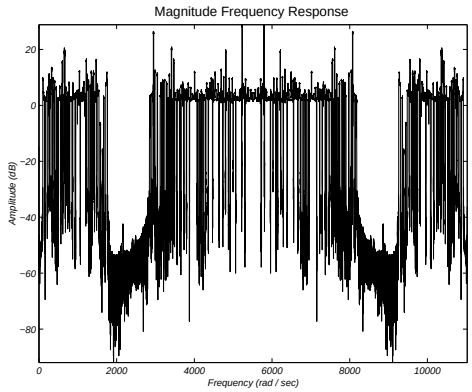
given by:

$$\hat{\phi}_k(\omega) \approx \psi_0 + \psi_1\,\omega + \psi_2\,\omega^2 \tag{9.41}$$

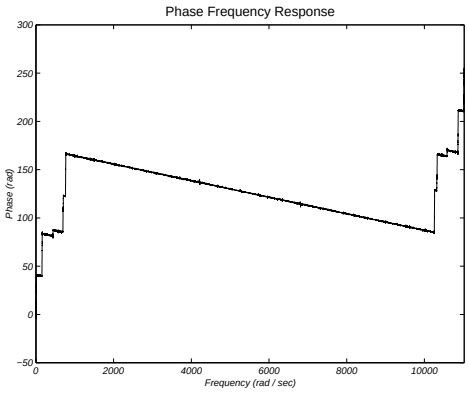
The coefficients ψ_i, $i \in \{0, 1, 2\}$ can be estimated using least-squares. Figure 9.16 shows the effect of ignoring $\phi_k(\omega)$: The equalised impulse response when $\phi_k(\omega)$ is not accounted for is shown in Figure 9.16(a). Compared with Figure 9.13(c), the equalised response is much longer, and thus no longer accurately reflects an impulse. Acoustic listening tests, in which a clean speech signal is filtered by each of the equalised responses in Figures 9.13(c) and 9.16(a) respectively, indicate that the speech is heavily distorted in the latter case when phase is not modelled. The magnitude response in Figure 9.16(b) indicates that the estimation of the gain terms in each subband is effected by ignoring the phase term. Finally, the phase responses of the equalised impulse response in the two cases are show in Figures 9.16(c) and 9.16(d). It can be seen that when the additional phase term not modelled, the phase response is far from linear, and therefore introduces distortion. Hence, it is important to model $\phi_k(\omega)$ when using the subband autoregressive model.



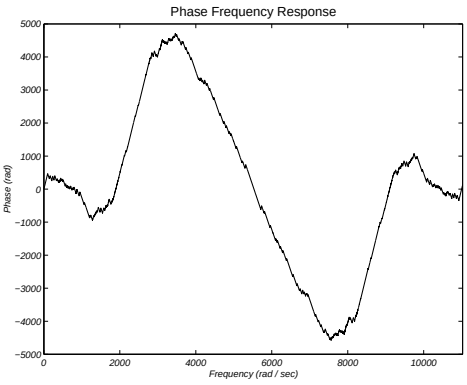
(a) Inverse impulse response when the additional phase term is not modelled.



(b) Magnitude response of impulse response in Figure 9.16(a).



(c) Equalised phase response when additional phase term is modelled using equation (9.41).



(d) Equalised phase response when additional phase term is ignored.

Figure 9.16: *Effect of ignoring additional phase term* $\phi_k(\omega)$.

9.8 Chapter Summary and Discussion: Restoration of Speech in an Acoustic Environment

Real room acoustic impulse responses modelled by an IIR filter require very large model orders. A problem with modelling a signal with complicated structure by an all-pole spectrum, is that the model must simultaneously fit the entire frequency range, even though the model may fit some regions of this frequency space better than others. This suggests it may be better to model a particular frequency band of

the filters spectrum by an all-pole model, which can often result in low model orders *for that frequency band* and, therefore, within this frequency band yield improved parameter estimation. In other words, the modelling of different frequency bands has been *decoupled* and, therefore, it is proposed that this will lead to a better model fit.

The techniques presented in this chapter should facilitate the restoration of speech in acoustic environments by using the subband autoregressive model presented in equation (9.22) of section §9.2.2: section §9.7 demonstrated that the subband model accurately represents typical acoustic impulse responses. Section §9.5 showed that the spectral-temporal model can accurately estimate the all-pole distortion filter, $\mathscr{A}$, in subbands, when driven by a nonstationary source. Thus, the foundations for the general approach for blind deconvolution of reverberant speech in real acoustic environments are laid. For nonminimum-phase acoustic impulse responses, subbands which possess minimum-phase characteristics can be inverted. Hence, in the case of a nonminimum-phase response, where a causal inverse does not exist, methods for detecting and equalising the minimum-phase subbands should be developed: the approaches in [384–386] may be of use.

Unfortunately, a problem which need to be addressed, and should be tackled in future work is as follows: If the length of the data blocks are small and the number of subbands is large, then the problem can become degenerate, since the system is considerably over parameterised. However, using few subbands means that the acoustic impulse response is under modelled, and thus restoration is poor. Nevertheless, this approach shows promising results, and this technique is the foundation for allowing high-order channel models to be estimated, which is simply not possible using 'full-band' techniques.

10

Part III Conclusions: Blind Deconvolution

SINGLE channel blind deconvolution is introduced in Chapter 7, and models a degraded observed signal, $x(t)$, as the convolution of the source signal, $s(t)$, with a distortion operator, $\mathscr{A}$. The distortion operator could, for example, represent the acoustical properties of a room, the effect of *multipath propagation* in the reception of radio signals, or a non-impulsive excitation in seismic applications. The task is to estimate the source signal *or* the distortion operator, given only the observed signal. The problem is under-constrained and *can only* be solved by incorporating varying degrees of *a priori* knowledge regarding $s(t)$ and $\mathscr{A}$. The general distinguishing features of the signal $s(t)$ and the operator $\mathscr{A}$ *must* be known, otherwise it becomes impossible to assess when the algorithm is performing correctly. An important characteristic of blind deconvolution is that the source signal and impulse response of the distortion operator must be *irreducible* for unambiguous deconvolution. When many linear systems are considered stationary, they become *reducible*, and *blind deconvolution* is impossible. However, if, in fact, $s(t)$ or $\mathscr{A}$ are both stationary and *reducible* over a short period of time, but possess different rates

of *global nonstationarity* then, $s(t)$ and $\mathscr{A}$ are no longer *globally reducible*, are thus irreducible and, therefore, blind deconvolution is possible if this nonstationarity is taken into account. The principle of *nonstationarity* is used to produce superior results for these problems and this is investigated with application to acoustic dereverberation, a problem which arises in the applications discussed in Chapter 1.

In Chapter 8 the single channel blind deconvolution problem has been tackled by modelling the source signal as a block stationary AR process, and the distortion operator as an IIR filter. The Bayesian paradigm has been introduced as a means of parameter estimation, and posterior densities for the distortion filter parameters conditional on the observed data were derived. An interpretation of the mechanism that allows the blind deconvolution to be performed is discussed in section §8.2.1. A simple example of restoring speech recorded through a gramophone horn is given in section §8.6, where a MMAP estimate of the filter parameters is used. The important questions of selecting model order and block length (or changepoint locations) have been investigated in sections §8.3 and §8.5. It is observed that by utilising the nonstationarity of the system, less specific knowledge regarding the model of the source signal is required. As long as the model of the source signal is nonstationary, the stationary component of the system can be estimated.

Real room acoustic impulse responses modelled by an IIR filter require very large model orders. A problem with modelling a signal with complicated structure by an all-pole spectrum, is that the model must simultaneously fit the entire frequency range, even though the model may fit some regions of this frequency space better than others. This suggests it may be better to model a particular frequency band of the filters spectrum by an all-pole model, which can often result in low model orders *for that frequency band* and, therefore, within this frequency band yield improved parameter estimation. In other words, the modelling of different frequency bands has been *decoupled* and, therefore, it is proposed that this will lead to a better model fit.

The techniques presented in Chapter 9 facilitate the restoration of speech in acoustic environments by using the subband autoregressive model presented in equation (9.22) of section §9.2.2: section §9.7 demonstrated that the subband model accurately represents typical acoustic impulse responses. Section §9.5 showed that the spectral-temporal model can accurately estimate the all-pole distortion filter, $\mathscr{A}$, in subbands, when driven by a nonstationary source. Thus, the foundations for the general approach for blind deconvolution of reverberant speech in real acoustic environments are laid.

IV

Conclusions, Appendices and Bibliography

11

General Conclusions and Future Work

THE conclusions regarding the solutions to the signal separation and blind deconvolution problems have been outlined in Chapter 5 and Chapter 10 respectively and will not be repeated here. These final comments relate to nonstationarity in general, and the proposed future direction of this research.

11.1 Nonstationary Signal Processing

A considerable number of linear signal processing problems reduce to the fundamental tasks of *signal separation* and *deconvolution*. A large proportion of these problems are also *blind* in the sense that neither of the source signals are known, and this substantially increases the difficulty of the problem. The estimation of signals in the presence of noise is a fundamental problem which occurs in many areas of science and technology. Some particular application areas, in which it is required to estimate a speech signal in noise, are telecommunications, voice input to a computer in an office environment, audio processing, and forensic science. At present,

hearing aids and hands-free telephones cannot compensate for the distortion due to reverberation. Further, radio waves, in many respects, propagate in a similar way to sound and, consequently, waves transmitted from a mobile are blurred by reflections in built-up areas, a reverberation effect known as *multipath propagation*. Current mobile phones circumvent this problem by transmitting a training sequence which the receiver uses to estimate the channel distortion. However, this approach increases the bandwidth required by the telephone and, consequently, increases call charges. Solutions to these problems can be attempted using nonstationary signal processing: the research presented in this dissertation indicates that superior results are obtained by discarding the stationary assumption, and utilising nonstationarity.

Nonstationary modelling and improved source modelling work is of theoretical interest to the academic community. As well as in audio restoration, this work should find wider practical application in other model-based signal processing applications such as statistical data analysis, econometrics, biomedical data, environmental engineering, the physical sciences, and in particular speech and audio coding and communications channel equalisation.

11.2 Suggestions for Future Research

A general aim of this research was to advance the understanding and application of nonstationary signal processing, with a focus on the potential gains of utilising nonstationarity within a system. As such, continuing with this philosophy, the following lines of research are suggested:

- Demonstrate the advantage of using authentic nonstationary signal models in the modelling of nonstationary systems, and investigate methods for choosing the most appropriate nonstationary model.
- Develop a theoretical framework to determine how nonstationarity can best be utilised, and then, using appropriate nonstationary modelling, apply it to produce superior results both in existing problems solved using the stationarity assumption, and previously intractable problems, especially in the areas of speech enhancement, signal separation, and noise reduction.
- Investigate whether it is possible to explicitly quantify the *degree* of nonstationarity required to find solutions for a particular problem and, if so,

whether there is a trade-off between the complexity of the nonstationarity and the degree of prior information required in the model.

- Extend the nonparametric theoretical framework for single channel signal separation into a parametric framework for application where parametric signal models are known to be appropriate.

In some applications it is possible to postulate a parametric signal model and, under these conditions, it is generally the case that results obtained for signal estimation are significantly better than non-parametric methods. There are many issues to be resolved in parametric modelling of nonstationary processes, but two major ones are as follows. An ideal model would be capable of describing a nonstationary signal, for example, a speech signal, at all time instants. However, such a model would be overly complex and a more appropriate approach would be to define a set of models. For example, in speech one would require different models for each of the classes of speech sound, namely voiced, unvoiced, and so forth. Thus at each time instant it is necessary to estimate the signal using the 'most appropriate' model, and this requires both estimation and model selection. A consistent framework for achieving these objectives is the Bayesian statistical methodology, and this general approach should be investigated further.

The following two applications specifically discuss extensions to the work on signal separation and blind deconvolution presented in this dissertation.

11.2.1 Blind Image Restoration

As noted by Lagendijk *et al*. [202], 'perhaps the best approach to the restoration of noisy blurred images would be to prevent the degradation from occurring at all.' Unfortunately, in many applications these degradations cannot be avoided and a noisy blurred image is observed. These degradations are due to, for example, motion blur, out-of-focus blur, and atmospheric turbulence; Figure 11.1 shows typical examples of some image degradations. In some applications, such as astronomy, adaptive-optical systems may be used to compensate for blurring, although the high cost of such systems makes this impractical for some observational facilities [198]. In other situations a blurred image must be accepted as given: for example, a blurred picture of a unique event. According to human factor studies, 15 to 20 per cent of consumer pictures have detectable image-blur problems mainly

(a) Original image. (b) Out-of-focus image. (c) Motion blurred image.

Figure 11.1: *Original and degraded images.*

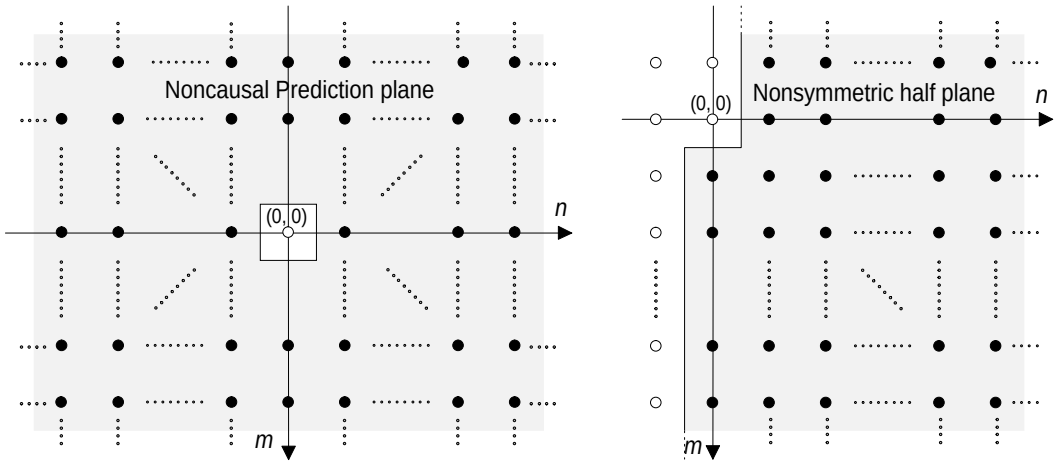
due to focus error, camera shake, and object motion [212]. Recent surveys of various blind image deconvolution approaches commonly used are given in Kundur and Hatzinakos [198,199] and Sezan and Tekalp [320]. This section proposes how the techniques presented in Chapter 8 could be generalised to the two-dimensional case, and thus to image restoration.

11.2.1.1 Source Image Model

In many image restoration techniques, the discrete original image is often represented by a class of models known as *spatial interaction models* [67, 167], which are characterised by the image property that the gray level at a lattice point is statistically dependent on those of its neighbours. An $M \times N$ image is described by a finite set of gray levels $\{s(m,n),\ (m,n) \in \mathscr{I}\}$, $\mathscr{I} = \{1,\ldots,M\}\times\{1,\ldots,N\} \subset \mathbb{Z}\times\mathbb{Z}$, where $s(m,n)$ is the gray level of the lattice point (m,n), and is often modelled by a zero-mean, homogeneous, discrete $M\times N$ random field which may be represented by the 2-D Markov image model [166, 167, 397, 398]:

$$\begin{aligned} s(m,n) &\triangleq -b(m,n) \star s(m,n) + e(m,n) \\ &= -\sum_{\substack{(p,q)\in\mathscr{W} \\ (p,q)\neq(0,0)}} b(p,q)\, s(m-p,n-q) + e(m,n), \quad (m,n) \in \mathscr{I} \end{aligned} \tag{11.1}$$

where $b(p,q)$, $(p,q) \in \mathscr{W}$, are the image model coefficients, $e(m,n)$ is the modelling error, which is a zero-mean homogeneous noise process, and $\mathscr{W} \subset \mathscr{I}_{-\{\emptyset\}} \triangleq$



(a) Noncausal Prediction Plane, $\mathscr{I}_{-\{\emptyset\}} \triangleq \mathscr{I} - \{\emptyset\}$ such that $(p, q) \in \mathscr{W} \Rightarrow (-p, -q) \in \mathscr{W}$.

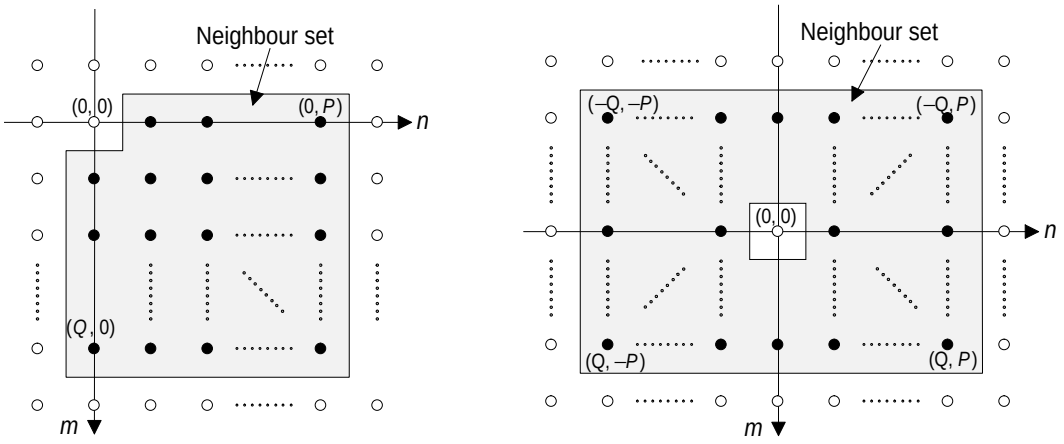
(b) Nonsymmetric Half Plane, $\mathscr{I}^+ \subset \mathscr{I}_{-\{\emptyset\}}$, such that $(m, n) \in \mathscr{I}^+$ and $(\hat{m}, \hat{n}) \in \mathscr{I}^+, \Rightarrow (m + \hat{m}, n + \hat{n}) \in \mathscr{I}^+, (-m, -n) \notin \mathscr{I}^+$.

Figure 11.2: *Prediction regions for Gaussian Markov random fields.*

$\mathscr{I} - \{\emptyset\}$ (see Figure 11.2(a)) is a set of points $\{(p, q) : (0, 0) \neq (p, q) \in \mathbb{Z} \times \mathbb{Z}\}$ called the *neighbour set* or the *neighbour region of support*, which are subsets of *prediction regions*, typical examples of which are shown in Figure 11.2 [124, 167]. The most widely used models which satisfy this form are the Gaussian Markov random field (GMRF) and *2-D autoregressive* models. Typical examples of neighbour sets used in GMRF models are shown in Figure 11.3. Due to its generality, the 2-D Markov image model defined in equation (11.1) has been of much interest in image processing [38, 66, 68, 166, 167, 397, 398] and, in particular, image restoration [39, 67, 198, 202, 203, 349, 351]. For reference, it is noted that Katsaggelos and Lay [183, 210] model the image as a general zero-mean Gaussian process, and use the expectation-maximization (EM) algorithm to estimate the image and blurring function.

11.2.1.2 Nonhomogeneous Image Model

Images inherently possess nonstationary spatial statistics and, consequently, the model coefficients in an image model will be spatially-variant. Boudaoud and Chaparro [48] proposed a nonstationary image model composed of a space-varying mean and a nonstationary, zero mean, 2-D autoregressive (2-D AR) random field; in each case the space varying parameters are parameterised using complete orthogo-



(a) Unilateral Neighbour Set, $\mathscr{W} \subset \mathscr{I}^{+}$, where $\mathscr{I}^{+}$ is shown in Figure 11.2(b).

(b) Bilateral Neighbour Set, $\mathscr{W} \subset \mathscr{I}_{-\{\emptyset\}}$, where $\mathscr{I}_{-\{\emptyset\}}$ is shown in Figure 11.2(a).

Figure 11.3: *Neighbour sets for Gaussian Markov random fields.*

nal 2-D basis functions in a directly analogous way to the basis-function approach discussed in section §7.5.2 and [63, 130, 131, 225]. Further approaches for dealing with the nonstationarity of images have been discussed by Hunt and Cannon [165], Jain [167], Strickland [338] and Tekalp *et al.* [350] among many others. Keeping the philosophy that utilising nonstationarity reduces the sensitivity of the source model, a block stationary 2-D AR model analogous to that discussed in section §7.6 for 1-D signals is proposed. The image is partitioned into C columns and R rows, giving CR subimages. A compact notation for equation (11.1) can be arrived at by *lexicographic* ordering of the image data, which maps a matrix to a column vector [166, 167, 202]. Using such notation, the error process for each subimage, $(i, j), i \in \mathscr{R} = \{1, \ldots, R\}$, $j \in \mathscr{C} = \{1, \ldots, C\}$, may be written as:

$$\mathbf{e}_{ij} = \mu_{ij} + \mathbf{s}_{ij} + \mathcal{S}_{ij}\, \mathbf{b}_{ij} \tag{11.2}$$

where $\mu_{ij} \triangleq \mu_{ij}\, \mathbf{1}$, with $\mathbf{1}$ as a column vector of 1's, is introduced as the mean of each subimage since, although the mean of the entire image is assumed zero, the same cannot be said of each subimage.

Although the GMRF model is intellectually appealing, since the actual model parameters provide a minimum mean-square estimate (MMSE) of the error signal, unless the neighbour set $\mathscr{W}$ is unilateral, the covariance function, $R_{ee}(m, n)$, for the

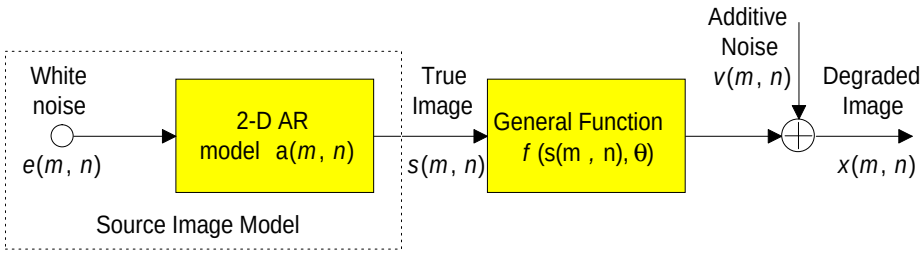


Figure 11.4: *Degraded image model for the output of a general system* $\mathrm{x}(\mathrm{m},\mathrm{n}) = \mathrm{f}(\mathrm{s}(\mathrm{m},\mathrm{n}),\boldsymbol{\theta})$ *driven by a 2-D autoregressive image model.*

error process, $\{\mathrm{e}(\mathrm{m},\mathrm{n})\}$, is dependent on the model parameters, b_{ij}, and, in general, the posterior distribution becomes intractable to analytic solution. Casual GMRF or casual 2-D AR processes do not employ all the image information available despite its successful use in image modelling. As such, in order to procure a closed-form expression for the posterior distribution, the image will be modelled by a noncausal AR (NCAR) process which, while intuitive and retaining intellectual attraction, has also been shown to successfully model textures [68].

With the image model in the form of equation (11.2), the posterior distribution for a general system driven by a nonhomogeneous image model can be derived in a similar manner to the approach discussed in section §7.7.3. A complete model for such a degraded image is shown in Figure 11.4 where, for completeness, additive observation noise has been included. However, the effectiveness of this model must be investigated in more detail.

11.2.2 Linear Time-Frequency Transforms

Alternative definitions of power spectra using time-frequency techniques were discussed in section §3.2.1, and a natural extension to the linear transform techniques discussed Chapter 3 is the reformulation of the results for linear time-frequency methods. The representation of a *continuous* time deterministic signal $\mathrm{x}(\mathrm{t})$, $\mathrm{t} \in \mathrm{T} \subset \mathbb{R}$ on an arbitrary *continuous joint* spectral domain $(\phi,\psi) \subset \mathbb{C} \times \mathbb{C}$ is $\mathrm{X}(\phi,\psi)$, could be defined by the general *joint* integral transform:

$$\mathrm{X}(\phi,\psi) = \int_{\mathrm{T}} \mathrm{x}(\mathrm{t})\, \mathrm{K}(\mathrm{t},\phi,\psi)\, \mathrm{d}\mathrm{t}\,, \quad \forall\,(\phi,\psi) \in (\Phi,\Psi) \tag{11.3a}$$

where $(\Phi \times \Psi)$ is the region in the signal space in which the representation $X(\phi, \psi)$ lies. The function $K(t, \phi, \psi)$ is called the *forward joint transform basis kernel*. Conversely, $X(\phi, \psi)$ may be represented on the time domain $t \in T \subset \mathbb{R}$ as $x(t)$;

$$x(t) = \iint_{\Phi\Psi} X(\phi, \psi)\, k(\phi, \psi, t)\, d\psi\, d\phi\,, \quad \forall t \in \mathscr{T} \tag{11.3b}$$

where $k(\phi, \psi, t)$ is called the *inverse joint transform kernel* of the *reciprocal joint basis kernel* of $K(t, \phi, \psi)$. Since this system is redundent, however, the functions $k(\phi, \psi, t)$ must be linearly dependent, where ϕ and ψ are considered as parameters, and therefore expressable in terms of a set of linearly independent basis functions $k(\lambda, t)$ such that:

$$k(\phi, \psi, t) = \int_{\Lambda} \varphi(\phi, \psi, \lambda)\, k(\lambda, t)\, d\lambda \tag{11.4}$$

and, consequently, $x(t)$ can be expressed in terms of linearly independent functions using equation (3.8b):

$$x(t) = \int_{\Lambda} X(\lambda)\, k(\lambda, t)\, d\lambda\,, \ \forall t \in T \tag{3.8b}$$

where

$$X(\lambda) = \iint_{\Phi\times\Psi} X(\phi, \psi)\, \varphi(\phi, \psi, \lambda)\, d\psi\, d\phi \tag{11.5}$$

It is assumed that the transformation is isomorphic so, for a given $x(t)$, there exists a unique $X(\lambda, \phi)$ and, conversely, for a given $X(\lambda, \phi)$ there exists a unique $x(t)$. It therefore follows that the joint-kernels $k(\lambda, \phi, t)$ and $K(t, \lambda, \phi)$ must satisfy

$$\delta(t - \tau) = \iint_{\Lambda\Phi} K(t, \lambda, \phi)\, k(\lambda, \phi, \tau)\, d\phi\, d\lambda, \quad \forall\, (t, \tau) \in \mathscr{T} \tag{11.6a}$$

$$\delta(\lambda - \hat{\lambda}, \phi - \hat{\phi}) = \int_{\mathscr{T}} k(\lambda, \phi, t)\, K(t, \hat{\lambda}, \hat{\phi})\, dt, \quad \forall\, (\lambda, \hat{\lambda}) \in \boldsymbol{\Lambda}, \forall\, (\phi, \hat{\phi}) \in \boldsymbol{\Phi} \tag{11.6b}$$

where $\boldsymbol{\Lambda} = \Lambda \times \Lambda$ and $\boldsymbol{\Phi} = \Phi \times \Phi$.

The theoretical results in Chapter 3 could be extended to deal with the transform defined by equation (11.3), although a complication occurs for redundant systems. Hence, care must be taken when extensing the concept of the ideal filter introduced in section §4.1.

A

Psychology of Hearing

The binaural hearing mechanism exploits the *spectral*, *temporal* and *spatial* characteristics of natural sounds to aid the identification and separation of various sound sources. The hearing mechanism achieves this using several techniques which can be compared with methods known in the field of statistical signal processing, namely,

1. Frequency selectivity (*cf.* Bandpass filtering)
2. Spatial-directionality, using
   (a) Temporal delay for low frequency signals (*cf.* Beam-forming)
   (b) Attenuation for high frequency signals (*cf.* Equalisation)
3. Spectral spatial directionality (*cf.* Beam-forming)

A full discussion of these mechanisms can be found in, for example, Lindsay and Norman [223], and this appendix briefly discusses the fundamental principles. The properties of the hearing mechanism provide a useful insight into the features that signal processing techniques might exploit in order to achieve blind deconvolution and blind signal separation.

A.1 Frequency Sensitivity

In the *cochlea*, inside the ear, there is a structure called the *basilar membrane*, each part of which acts as a narrow bandpass filter. This frequency selectivity is useful in the perception of different types of sounds, for example, to separate speech from music. The human ear can therefore reconstruct pitch and frequency information from a monaural source and, as such, monophonic recordings are often effective for single speakers. There is, of course, a spatial–amplitude ambiguity in mono recordings, since they do not convey any spatial information through acoustical cues.

A.2 Spatial Perception

The ability to localise sound sources is of immense importance in

1. determining the direction in which one should focus their attention,
2. estimating the approximate distance of the sound source from the ear, and
3. performing signal separation.

The latter is possible by analogy with beam-forming techniques: by looking at a source location, noise sources from other directions are attenuated due to:

1. the geometric positioning of the sound sources relative to the ears (see below), and
2. the ability for humans to correlate sound sources with visual cues [128].

Sound localisation using echoes and reflections is, for example, an invaluable feature for blind people to avoid obstacles. Most acoustic cues in sound localisation are based upon a comparison of the signals reaching the two ears, hence *binaural hearing*. A sound source is located using a slightly different mechanism for high frequency sounds and low frequency ones [223], as discussed below.

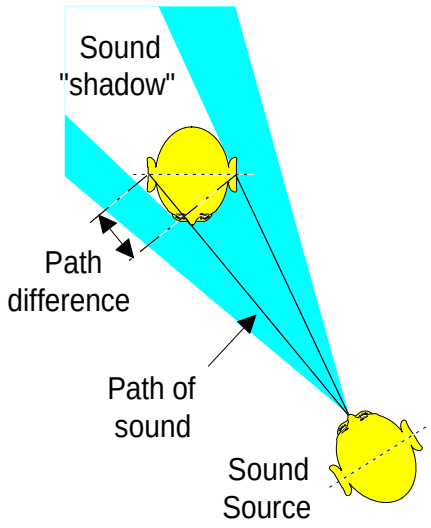


Figure A.1: *A sound source is located using different mechanisms for different frequency sounds. Low frequency components are located by the relative time difference between the sound arriving at each ear, whilst high frequency components are located by considering the attenuation between each ear.*

- The low frequency components of a signal bend, or refract, around the head with little attenuation. This can be demonstrated straightforwardly by referring to Figure A.1. Supposing the sound source is to the left and in front of the listener; the sound waves reach the left ear before the right one. The waves which arrive *tangentially* to the head will refract around to the right ear, and the time difference between the signal reaching each ear is related, perhaps non-trivially,[1] to the angle of arrival.

- The signal components that are of frequency greater than about 1.3 kHz impart a phase ambiguity between the ears, and it is not possible to accurately localise the sound source. This is, in part, the reason why one is never sure which direction the traditional high-frequency sirens used by the emergency services originate from. However, high frequency components *do not* refract around the head, and a sound shadow is created, which attenuates the signal between the ears. This attenuation then provides a key to the localisation of high frequency sources.

- Finally, the direction dependent filtering produced by the *pinna*, or outer ear, and upper body, especially the shoulders, is crucial for judgements of location in the vertical direction, and for front-back discrimination.

[1] The human brain is, of course, remarkably adept at learning this relationship given experience.

A.3 Impaired Hearing

Human hearing is a complex mechanism and there can be various causes at many different points for a loss of hearing. Damage of some organs in the ear can cause impairment or loss of hearing, and this dysfunction can be classified as:

1. *conductive loss*, which can be compensated for by using a simple amplifying hearing aid,
2. *functional* deafness, rectifiable only through surgery, or
3. *sensori-neural loss*, caused by damage in the basilar membrane of the cochlea.

Sensori-neural loss cannot be compensated for using a simple hearing aid, and may lead to a loss of frequency sensitivity, reduction in dynamic range, poor temporal resolution and a loss of spatial perception [56]. A person with this impairment is unable to separate two similar frequency components in a complex tone and, because of damaged auditory filters, background noise causes a significant decrease in intelligibility. Thus, for improved hearing, the hearing aid should increase the frequency selectivity, improve spatial perception (*reverberation cancellation*), and eliminate unwanted frequency components (*signal separation*).

B

Results in LTV System Theory

This appendix contains details of results and proofs used throughout Part II of this dissertation.

B.1 Karhunen-Loève Expansion

As discussed in section §3.2.3, the Karhunen-Loève (KL) expansion is a special case of the general integral transform. In this particular transform, the kernel, $k(t,\lambda)$, is self-reciprocal and, as such, is a set of orthonormal functions, given by equations (3.10):

$$\int_{\Lambda} k(t,\lambda)\, k^*(\tau,\lambda)\, d\lambda = \delta(t-\tau) \tag{B.1a}$$

$$\int_{T} k(t,\lambda)\, k^*(t,\hat{\lambda})\, dt = \delta(\lambda-\hat{\lambda}) \tag{B.1b}$$

A random process, $\mathbf{x}(t)$, can therefore be expanded into a series of the form

$$\mathbf{X}(\lambda) = \int_T \mathbf{x}(t)\, k^*(t,\lambda)\, dt\,, \quad \forall \lambda \in \Lambda \tag{B.2a}$$

$$\hat{\mathbf{x}}(t) = \int_\Lambda \mathbf{X}(\lambda)\, k(t,\lambda)\, d\lambda\,, \quad \forall t \in T \tag{B.2b}$$

Further, the orthogonal functions are the solution of the integral equation [286]:

$$\int_T R_{xx}(t,\tau)\, k(\tau,\lambda)\, d\tau = \phi(\lambda) k(t,\lambda) \tag{B.3a}$$

Thus, it can be shown that, as a consequence of the positive definite characteristic of the autocorrelation function, the functions $k(t,\lambda)$ satisfy the identity:

$$R_{xx}(t,t) = \int_\Lambda \phi(\lambda)\, |k(t,\lambda)|^2\, d\lambda \tag{B.3b}$$

A useful characteristic of the Karhunen-Loève transform is that the spectral components, $\mathbf{X}(\lambda)$, are uncorrelated as the following theorem describes:

Theorem 22 (Orthogonality of Karhunen-Loève Transform) *The representation* $\hat{\mathbf{x}}(t)$ *equals* $\mathbf{x}(t)$ *in the MS sense:*

$$E\left[|\mathbf{x}(t) - \hat{\mathbf{x}}(t)|^2\right] = 0 \tag{3.19}$$

and if the basis kernels, $k(t,\lambda)$*, are the eigenfunctions of equation* (B.3a)*, the spectral components* $\mathbf{X}(\lambda)$ *are uncorrelated:*

$$E\left[\mathbf{X}(\lambda_t)\mathbf{X}(\lambda_\tau)\right] = \phi(\lambda_t)\delta(\lambda_t - \lambda_\tau) \tag{B.4}$$

PROOF. Using equation (3.25b) and equation (B.3a), gives the identity:

$$E\left[\mathbf{X}(\lambda)\, \mathbf{x}^*(\tau)\right] = \int_T R_{xx}(\alpha,\tau)\; k^*(\alpha,\lambda)\, d\alpha = \phi(\lambda) k^*(\tau,\lambda)$$

and, using the spectral representation for $\mathbf{X}(\lambda)$, then:

$$\begin{aligned} E\left[\mathbf{X}(\lambda_t)\, \mathbf{X}^*(\lambda_\tau)\right] &= \int_T E\left[\mathbf{X}(\lambda_t)\, \mathbf{x}^*(\alpha)\right]\, k(\alpha,\lambda_\tau)\, d\alpha \\ &= \int_T \phi(\lambda_t)\, k^*(\alpha,\lambda_t)\, k(\alpha,\lambda_\tau)\, d\alpha = \phi(\lambda_t) \int_T k^*(\alpha,\lambda_t)\, k(\alpha,\lambda_\tau)\, d\alpha \end{aligned}$$

and, hence, using equation (B.1a) gives equation (B.4). □

In the stationary case, the KL transform reduces to the Fourier transform [286,335]:

Theorem 23 (Karhunen-Loève Expansion for Stationary Signals) *The Karhunen-Loève (KL) transform defined for all time, for a WSS process, is equivalent to the Fourier transform:* i.e.

$$k(t,\lambda) = e^{j\lambda t} \tag{B.5}$$

PROOF. The orthogonal kernel satisfies (B.3a) and so, since $\mathbf{x}(t)$ is a WSS process,

$$\int_T R_{xx}(t-\tau)\, k(\tau,\lambda)\, d\tau = \phi(\lambda) k(t,\lambda) \tag{B.6a}$$

The eigenfunctions of this equation may be shown to be $k(t,\lambda) = e^{j\lambda t}$ [335] and can be verified by substitution:

$$\int_T R_{xx}(t-\tau)\, e^{j\lambda\tau}\, d\tau = e^{j\lambda t} \int_T R_{xx}(t)\, e^{-j\lambda\tau}\, d\tau = \phi(\lambda) e^{j\lambda t} \tag{B.6b}$$

and, hence, the corresponding eigenvalue is the power spectrum at frequency λ. □

B.2 Separable Impulse Response

The separable form of an impulse response, $h(t,\tau)$, as expressed in Theorem 8 for LTV differential equations, is well known as discussed in, for example, [417]. D'Angelo [84, Chapter 2] gives a proof of this result avoiding the use of transform methods. This theorem, and an outline of its proof, are given below.

B.2.1 Finite-Order Linear Differential Equation

Theorem 24 (Separable Impulse Response) *All impulse responses of linear systems characterised by the scalar linear time-varying differential equation:*

$$\mathscr{L}\, x(t) = \alpha_n\, a_n(t) \frac{d^n y(t)}{dt^n} + \alpha_{n-1}\, a_{n-1}(t) \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + \alpha_0\, a_0(t)\, y(t) = u(t) \tag{3.61}$$

are separable, i.e.,
$$h(t,\tau) = \sum_{i=1}^{n} p_i(t)\, q_i(\tau), \quad t \geq \tau \tag{B.7}$$

Furthermore, $\{p_i(t)\}$ *are the basis functions of the homogeneous differential equation* $\mathscr{L}\, x(t) = 0$, *and* $\{q_i(t)\}$ *are the basis functions of the adjoint homogeneous differential equation* $\mathscr{L}^*\, x(t) = 0$.

PROOF. Consider a linear differential equation:

$$\mathscr{L}\, x(t) = \alpha_n\, a_n(t)\, \frac{d^n y(t)}{dt^n} + \alpha_{n-1}\, a_{n-1}(t)\, \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + \alpha_0\, a_0(t)\, y(t) = u(t) \tag{3.61}$$

This can be solved using a state-space approach. Define n new variables given by:

$$x_i(t) \equiv \frac{d^{(i-1)} y(t)}{dt^{(i-1)}} \quad i \in \{2, \ldots, n\}$$

Hence,
$$\frac{dx_i(t)}{dt} = \frac{d^i y(t)}{dt^i} = x_{i+1}(t) \quad i \in \{2, 3, \ldots, n-1\}$$

and
$$x_1(t) = y(t)$$

$$\frac{dx_n(t)}{dt} = -\frac{\alpha_0\, a_0(t)}{\alpha_n\, a_n(t)} x_1(t) - \cdots - \frac{\alpha_{n-1}\, a_{n-1}(t)}{\alpha_n\, a_n(t)} x_n(t) + \frac{u(t)}{\alpha_n\, a_n(t)}$$

This can now be written in the matrix form:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A(t)\, \mathbf{x}(t) + \mathbf{b}(t)\, u(t) \\ y(t) &= \mathbf{c}^T(t)\, \mathbf{x}(t) \end{aligned} \tag{B.8a}$$

where the time-varying matrix, $A(t)$, is given by:

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\frac{\alpha_0\, a_0(t)}{\alpha_n\, a_n(t)} & -\frac{\alpha_1\, a_1(t)}{\alpha_n\, a_n(t)} & -\frac{\alpha_2\, a_2(t)}{\alpha_n\, a_n(t)} & \cdots & -\frac{\alpha_{n-1}\, a_{n-1}(t)}{\alpha_n\, a_n(t)} \end{bmatrix}$$

and the time-varying vectors $\mathbf{b}(t)$, $\mathbf{c}(t)$, and $\mathbf{x}(t)$ are given as:

$$\mathbf{b}(t) = \begin{bmatrix} 0 & \cdots & 0 & \frac{1}{\alpha_n\, a_n(t)} \end{bmatrix}^T, \quad \mathbf{c}(t) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) & \cdots & x_n(t) \end{bmatrix}^T$$

or, defining $\mathbf{f}(t) = \mathbf{b}(t)u(t)$, then

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + \mathbf{f}(t) \tag{B.8b}$$

Solutions to this differential equation are found by introducing the *adjoint equation* defined by:

$$\dot{\mathbf{x}}^*(t) = -A^T(t)\,\mathbf{x}^*(t) \tag{B.8c}$$

It can be shown that for a given initial condition $\mathbf{x}_i^*(t_0) \equiv \mathbf{x}_{i0}^*$, a unique solution exists to equation (B.8c). A basis for this adjoint system may be formed by taking n distinct initial conditions, and assuming that these initial conditions are chosen such that there are n independent solutions to the adjoint equation, $\mathbf{x}_i^*(t)$, $i = 1, \ldots, n$. Consider the scalar differential equation:

$$\frac{d}{dt}[\mathbf{x}_i^{*T}(t)\mathbf{x}(t)] = \mathbf{x}_i^{*T}(t)\,\mathbf{f}(t) \tag{B.9a}$$

On expansion, this gives:

$$\mathbf{x}_i^{*T}(t)\,\mathbf{f}(t) = \mathbf{x}_i^{*T}(t)\,\dot{\mathbf{x}}(t) + \mathbf{x}^T(t)\,\dot{\mathbf{x}}^*(t)$$

and, using equation (B.8c),

$$\mathbf{x}_i^{*T}(t)\,\mathbf{f}(t) = \mathbf{x}_i^{*T}(t)\,\dot{\mathbf{x}}(t) - \mathbf{x}^{*T}(t)\,A(t)\,\mathbf{x}(t) \tag{B.9b}$$

$$= \mathbf{x}_i^{*T}(t)\,\{\dot{\mathbf{x}}(t) - A(t)\,\mathbf{x}(t)\} \tag{B.9c}$$

Thus, this solution satisfies the original differential equation (B.8b). Combining these solutions, and defining the matrix:

$$\Phi(t) \equiv \begin{bmatrix} \mathbf{x}_1^{*T}(t) \\ \mathbf{x}_2^{*T}(t) \\ \vdots \\ \mathbf{x}_n^{*T}(t) \end{bmatrix}$$

equation (B.9a) may be written as

$$\frac{d}{dt}[\Phi(t)\mathbf{x}(t)] = \Phi(t)\mathbf{f}(t) \tag{B.10}$$

Integrating with respect to (w. r. t.) time,

$$\Phi(t)\mathbf{x}(t) = \Phi(t_0)\mathbf{x}_0 + \int_{t_0}^{t} \Phi(\tau)\mathbf{f}(\tau)\, d\tau \tag{B.11a}$$

gives

$$\mathbf{x}(t) = \Phi^{-1}(t)\Phi(t_0)\mathbf{x}_0 + \int_{t_0}^{t} \Phi^{-1}(t)\Phi(\tau)\mathbf{f}(\tau)\, d\tau \tag{B.11b}$$

Noting $y(t) = \mathbf{c}^T(t)\mathbf{x}(t)$, the solution of the original differential equation is:

$$y(t) = \mathbf{c}^T(t)\Phi^{-1}(t)\Phi(t_0)\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{c}^T(t)\Phi^{-1}(t)\Phi(\tau)\mathbf{b}(\tau)u(\tau)\, d\tau \tag{B.11c}$$

Comparing this with the superposition integral in equation (3.35), the impulse response of the system is given by:

$$h(t,\tau) = \mathbf{c}^T(t)\,\Phi^{-1}(t)\Phi(\tau)\,\mathbf{b}(\tau) = \mathbf{p}^T(t)\mathbf{q}(\tau)$$

where

$$\mathbf{p}^T(t) = \mathbf{c}^T(t)\Phi^{-1}(t)$$

$$\mathbf{q}(t) = \Phi(\tau)\mathbf{b}(\tau)$$

On expansion, this is in the form of equation (B.7), as required. □

B.2.2 Relating Finite and Infinite Dimensional Cases

Theorems 8 and 24 can be linked by setting in equation (3.41),

$$K(t,\lambda) = \sum_i q_i(\tau)\,\delta(\lambda - \lambda_i) \quad \text{and} \quad \hat{k}(\lambda_i, t) = p_i(t)$$

which would thus give:

$$\begin{aligned} h(t,\tau) &= \int_{\Lambda_H} \hat{k}(\lambda, t)\left(\sum_i q_i(\tau)\,\delta(\lambda - \lambda_i)\right) d\lambda \\ &= \sum_i q_i(\tau) \int_{\Lambda_H} \hat{k}(\lambda, t)\,\delta(\lambda - \lambda_i)\, d\lambda \\ &= \sum_i q_i(\tau)\, k(\lambda_i, t) = \sum_i p_i(t)\, q_i(\tau) \end{aligned}$$

as required. □

C

Wiener Filter Equations and Solutions

This appendix contains results pertaining to the Wiener-Hopf filter (WHF).

C.1 Derivation of Wiener-Hopf Filter

Proof (of Theorem 1). Given the estimate, $\hat{\mathbf{d}}(t)$, of the desired signal, $\mathbf{d}(t)$, as:

$$\hat{\mathbf{d}}(t) = \int_T h(t, \alpha)\, \mathbf{x}(\alpha)\, d\alpha \tag{3.4}$$

the error between the actual desired signal and its estimate is given by:

$$\varepsilon(t) = \mathbf{d}(t) - \hat{\mathbf{d}}(t) = \mathbf{d}(t) - \int_T h(t, \alpha)\, \mathbf{x}(\alpha)\, d\alpha \tag{C.1a}$$

The impulse response is found by minimising the mean squared error (MSE), which is given by:

$$\mathrm{E}\left[\varepsilon(t)\varepsilon^*(t)\right] = \mathrm{E}\left[\left(\mathbf{d}(t) - \int_T h(t,\alpha)\,\mathbf{x}(\alpha)\,d\alpha\right)\left(\mathbf{d}^*(t) - \int_T h^*(t,\alpha)\,\mathbf{x}^*(\alpha)\,d\alpha\right)\right] \tag{C.1b}$$

Differentiating w. r. t. $h(\tau,\hat{\tau})$, noting the expectation operator is linear:

$$\begin{aligned}\frac{\partial \mathrm{E}\left[\varepsilon^2(t)\right]}{\partial h(\tau,\hat{\tau})} &= \mathrm{E}\left[-2\varepsilon(t)\frac{\partial}{\partial h(\tau,\hat{\tau})}\left(\int_T h(t,\alpha)\,\mathbf{x}(\alpha)\,d\alpha\right)\right] \\ &= -2\mathrm{E}\left[\varepsilon(\tau)\mathbf{x}(\hat{\tau})\right] = -2\mathrm{E}\left[\mathbf{d}(\tau)\mathbf{x}(\hat{\tau}) - \int_T h(\tau,\alpha)\,\mathbf{x}(\alpha)\,\mathbf{x}(\hat{\tau})\,d\alpha\right]\end{aligned} \tag{C.1c}$$

Equating this expression to zero, gives the nonstationary Wiener-Hopf equation:

$$R_{dx}(\tau,\hat{\tau}) = \int_T h(\tau,\alpha) R_{xx}(\alpha,\hat{\tau})\,d\alpha \tag{3.5a}$$

Now, consider again equation (C.1b); expanding using equation (C.1c):

$$\begin{aligned}\mathrm{E}\left[\varepsilon^2(t)\right] &= \mathrm{E}\left[\left(\mathbf{d}(t) - \int_T h(t,\alpha)\,\mathbf{x}(\alpha)\,d\alpha\right)\varepsilon^*(t)\right] \\ &= \mathrm{E}\left[\mathbf{d}(t)\varepsilon^*(t)\right] = \mathrm{E}\left[\mathbf{d}^2(t)\right] - \int_T h(t,\alpha)\,\mathrm{E}\left[\mathbf{d}(t)\mathbf{x}^*(\alpha)\right]\,d\alpha \\ &= R_{dd}(t,t) - \int_T h(t,\alpha) R_{dx}(t,\alpha)\,d\alpha\end{aligned} \tag{3.5b}$$

which is equation (3.5b). In the case when the MSE goes to zero, from equation (3.5b), the constraint in equation (3.5c) is obtained. □

C.2 Discrete Wiener Filter Equations

Theorem 25 (Discrete Nonstationary Wiener Filter) *The nonstationary Wiener-Hopf filter equations in finite discrete time are:*

$$R_{dx}(q,\hat{q}) = \sum_{m\in\mathcal{N}} h(q,m) R_{xx}(m,\hat{q}) \qquad \forall q \in N,\ \forall \hat{q} \in \mathcal{N} \tag{C.2a}$$

with the MSE given by:

$$\sigma^2(n) = R_{dd}(n,n) - \sum_{m\in\mathcal{N}} h(n,m) R_{dx}(n,m), \quad \forall n \in N \tag{C.2b}$$

For 'perfect' signal separation, $\{\sigma^2(n) = 0, \forall n \in N\}$*, and thus the autocorrelation functions must satisfy:*

$$R_{dd}(n,n) = \sum_{m\in\mathcal{N}} h(n,m) R_{dx}(n,m) \tag{C.2c}$$

PROOF. The proof of this theorem is essentially identical to that of Theorem 1. □

C.3 General Spectral Solution of Stationary Wiener-Hopf Filter

Bode and Shannon [43] give an excellent introduction to the solutions of the stationary Wiener filter, and their paper contains further results in addition to those given below. The following argument is not intended to be rigorous, but merely to give some insight into the form of the solution. The autocorrelation function of the stationary Wiener-Hopf filter, Theorem 2, can be represented by the Fourier transform. Hence, taking Fourier transforms of the stationary Wiener filter equation

$$R_{dx}(\tau) = \int_T h(\alpha)\, R_{xx}(\tau - \alpha)\, d\alpha, \quad \forall \tau \in T \tag{C.3a}$$

gives the analytic equation:

$$\mathcal{P}_{dx}(\omega) = H(\omega)\, \mathcal{P}_{xx}(\omega) \tag{C.3b}$$

The variance of the error for the stationary Wiener-Hopf filter is given by:

$$\sigma^2(t) = R_{dd}(0) - \int_{\mathcal{T}} h(\alpha)\, R_{dx}(\alpha)\, d\alpha, \quad \forall t \in \mathcal{T} \tag{C.3c}$$

which, in the case of perfect signal separation, reduces to

$$R_{dd}(0) = \int_{\mathscr{T}} h(\alpha)\, R_{dx}(\alpha)\, d\alpha, \quad \forall t \in \mathscr{T} \tag{C.3d}$$

Substituting the Fourier decomposition of $h(t)$ into equation (C.3d) gives, after some rearrangement,

$$R_{dd}(0) = \frac{1}{2\pi}\int_{\Omega} H(\omega)\, \mathcal{P}_{dx}(-\omega)\, d\omega \tag{C.4a}$$

Moreover, by definition of the autocorrelation function (see equation (3.6)), then:

$$R_{dd}(0) \equiv \frac{1}{2\pi}\int_{\Omega} \mathcal{P}_{dd}(\omega)\, d\omega \tag{C.4b}$$

then, comparing equation (C.4a) with (C.4b), a solution for perfect separation is (using equation (C.3b)):

$$\left.\begin{aligned} \mathcal{P}_{dd}(\omega) &= H(\omega)\, \mathcal{P}_{dx}(-\omega) \\ \mathcal{P}_{dx}(\omega) &= H(\omega)\, \mathcal{P}_{xx}(\omega) \end{aligned}\right\} \quad \forall \omega \in \Omega \tag{C.5a}$$

Consider the independent additive case: taking Fourier transforms of (3.3b) on page 37, gives:

$$\left.\begin{aligned} \mathcal{P}_{xx}(\omega) &= \mathcal{P}_{dd}(\omega) + \mathcal{P}_{nn}(\omega) \\ \mathcal{P}_{dx}(\omega) &= \mathcal{P}_{dd}(\omega) \end{aligned}\right\} \quad \forall \omega \in \Omega \tag{C.5b}$$

Hence, from (C.5a), assuming $\mathbf{d}(t)$ is real such that $\mathcal{P}_{dx}(\omega) = \mathcal{P}_{dx}(-\omega)$,

$$\begin{aligned} \mathcal{P}_{dd}(\omega) &= H(\lambda)\, \mathcal{P}_{dd}(\omega) \\ \mathcal{P}_{dd}(\omega) &= H(\lambda)\, [\mathcal{P}_{dd}(\omega) + \mathcal{P}_{nn}(\omega)] \end{aligned} \tag{C.5c}$$

Ergo, using the definition in section §4.1.3 and, in particular, equation (4.5) on page 83, perfect separation is possible if the two signals are disjoint in the Fourier domain, and $h(t)$ is an ideal filter.

C.4 Derivation of General Spectral Solution of Nonstationary Wiener-Hopf Filter

This section contains a proof of sufficiency of Theorem 18. A necessary condition for a solution is not presented, inasmuch as there may be solutions for signal structures other than that defined in equation (3.18).

PROOF. Consider taking the general spectral transform of the Wiener-Hopf filter equation (3.5a) using equation (3.24a) in Definition 4 on page 48:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint_{\mathcal{T}\mathcal{T}\mathcal{T}} h(t, \alpha)\, R_{xx}(\alpha, \tau)\; K(t, \lambda_t)\, K^*(\tau, \lambda_\tau)\, d\alpha\, dt\, d\tau$$

substituting equation (3.38a) on page 55 gives ($\forall \lambda_t,\ \lambda_\tau \in \Lambda$),

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint_{\mathcal{T}\mathcal{T}\mathcal{T}} \iint_{\Lambda\Lambda} H(\lambda_1, \lambda_2)\, R_{xx}(\alpha, \tau)\; k(\lambda_1, t) \times K(\alpha, \lambda_2)\, K(t, \lambda_t)\, K^*(\tau, \lambda_\tau)\, d\lambda_1\, d\lambda_2\, d\alpha\, dt\, d\tau$$

and rearranging the integral w. r. t. t gives:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint_{\mathcal{T}\mathcal{T}\Lambda\Lambda} H(\lambda_1, \lambda_2)\, R_{xx}(\alpha, \tau)\; K(\alpha, \lambda_2) \times K^*(\tau, \lambda_\tau) \int_{\mathcal{T}} k(\lambda_1, t)\, K(t, \lambda_t)\, dt\, d\lambda_1\, d\lambda_2\, d\alpha\, d\tau$$

Thus, simplifying using equation (3.10b):

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint_{\Lambda\mathcal{T}\mathcal{T}\Lambda} H(\lambda_1, \lambda_2)\, R_{xx}(\alpha, \tau)\; K(\alpha, \lambda_2)\, K^*(\tau, \lambda_\tau)\, \delta(\lambda_1 - \lambda_t)\, d\lambda_1\, d\alpha\, d\tau\, d\lambda_2$$

which, using the sifting property (3.9), and the definition for the spectral decomposition, yields:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \int_\Lambda H(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda \quad \forall \lambda_t, \lambda_\tau \in \Lambda \tag{C.6a}$$

Finally, assumming equation (4.23a) is the solution of the Wiener-Hopf equations, then equation (C.6a) may be written in the form:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \int_{\Lambda_d} \delta(\lambda_t - \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda + \int_{\Lambda_o} H_0(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda \qquad \text{(C.6b)}$$

Noting that $\Lambda_d \oplus \Lambda_n = \{\emptyset\}$, $\Lambda_d \oplus \Lambda_0 = \{\emptyset\}$ then, using the sifting property (3.9), the first term in equation (C.6b) is equivalent to $\mathcal{P}_{xx}(\lambda_t, \lambda_\tau)$ if $\lambda_t \in \Lambda_d$, and zero otherwise. Moreoever, if $\lambda_t \in \Lambda_d$ then, since $H_0(\lambda_1, \lambda_2) = 0$ if $(\lambda_1, \lambda_2) \notin \boldsymbol{\Lambda_o}$, the second term is zero. However, if $\lambda_t \in \Lambda_0$, the first term is zero, while the second term is nonzero. Hence, using the relationships in Theorem 6 and noting

$$P_{dx}(\lambda_t, \lambda_\tau) = \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) + \mathcal{P}_{dn}(\lambda_t, \lambda_\tau)$$

it follows, after some slight rearrangement, that:

$$\mathcal{P}_{dd}(\lambda_t, \lambda_\tau) = \begin{cases} \mathcal{P}_{xx}(\lambda_t, \lambda_\tau) \equiv \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) & \text{for } (\lambda_t, \lambda_\tau) \in \boldsymbol{\Lambda_d}, \\ \displaystyle\int_{\Lambda_o} H_0(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda - \mathcal{P}_{dn}(\lambda_t, \lambda_\tau) & \text{for } (\lambda_t, \lambda_\tau) \in \boldsymbol{\Lambda_o}, \\ 0 & \text{elsewhere} \end{cases} \qquad \text{(C.6c)}$$

where it has been noted that if $\lambda_t \notin \{\Lambda_0 \cap \Lambda_d\}$, then $\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = 0 \Rightarrow \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) = -\mathcal{P}_{dn}(\lambda_t, \lambda_\tau) = 0$ since $d(t)$ and $n(t)$ only have spectral components which overlap in the region Λ_0. This proves the first part of the theorem. Since $\mathbf{d}(t)$ is a real process, the mean squared error from (3.5b) may be written as:

$$\begin{aligned} \sigma^2(t) &= R_{dd}(t,t) - \int_{\mathscr{T}} h(t,\alpha)\, R_{dx}(t,\alpha)\, d\alpha \\ &\equiv R_{dd}(t,t) - \int_{\mathscr{T}} h(t,\alpha)\, R_{dx}(\alpha,t)\, d\alpha \end{aligned} \qquad \text{(3.5b)}$$

Writing equation (4.22a) in the expanded form

$$\begin{aligned} R_{dd}(t,\tau) = &\iint_{\Lambda_d \Lambda_d} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau)\, k(\lambda_t, t)\, k^*(\lambda_\tau, \tau)\, d\lambda_t\, d\lambda_\tau \\ &+ \iint_{\Lambda_o \Lambda_o} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau)\, k(\lambda_t, t)\, k^*(\lambda_\tau, \tau)\, d\lambda_t\, d\lambda_\tau \end{aligned}$$

and substituting this, along with the expanded expression in (4.23c):

$$h(t,\tau) = \int_{\Lambda_d} k(\lambda, t)\, K(\tau, \lambda)\, d\lambda + \iint_{\Lambda_0} H_0(\lambda, \hat{\lambda})\, k(\lambda, t)\, K(\tau, \hat{\lambda})\, d\lambda\, d\hat{\lambda} \tag{4.23c}$$

into (3.5b):

$$\sigma^2(t) = R_{dd}(t,t) - \int_{\mathscr{T}} h(t,\alpha)\, R_{dx}(t,\alpha)\, d\alpha, \quad \forall t \in T \tag{3.5b}$$

gives, after use of equation (3.10b) (on page 42) and equation (4.22c) (on page 91),

$$\begin{aligned}
\sigma^2(t) = {} & \iint_{\Lambda_d\Lambda_d} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau)\, k(\lambda_t, t)\, k^*(\lambda_\tau, t)\, d\lambda_t\, d\lambda_\tau + \iint_{\Lambda_0\Lambda_0} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau)\, k(\lambda_t, t)\, k^*(\lambda_\tau, t)\, d\lambda_t\, d\lambda_\tau \\
& - \iint_{\Lambda_0\Lambda_0} (\cdots) \int_{\Lambda_d} k(\lambda, t)\, \delta(\lambda - \lambda_1)\, d\lambda\, d\lambda_1\, d\lambda_2 \\
& - \iint_{\Lambda_d\Lambda_d} \mathcal{P}_{dd}(\lambda_1, \lambda_2)\, k^*(\lambda_2, t) \int_{\Lambda_d} k(\lambda, t)\, \delta(\lambda - \lambda_1)\, d\lambda\, d\lambda_1\, d\lambda_2 \\
& - \iiint_{\Lambda_0\Lambda_d\Lambda_d} (\cdots) \int_{\Lambda_0} H_0(\hat{\lambda}_1, \hat{\lambda}_2)\, \delta(\hat{\lambda}_2 - \lambda_1)\, d\hat{\lambda}_2\, d\lambda_2\, d\lambda_1\, d\hat{\lambda}_1 \\
& \quad - \iiint_{\Lambda_0\Lambda_0\Lambda_0} \mathcal{P}_{dx}(\lambda_1, \lambda_2)\, k(\hat{\lambda}_1, t)\, k^*(\lambda_2, t) \int_{\Lambda_0} H_0(\hat{\lambda}_1, \hat{\lambda}_2)\, \delta(\hat{\lambda}_2 - \lambda_1)\, d\hat{\lambda}_2\, d\lambda_2\, d\lambda_1\, d\hat{\lambda}_1
\end{aligned}$$

where the dots $(\cdots)$ in the third integral represents $\mathcal{P}_{dx}(\lambda_1, \lambda_2)\, k^*(\lambda_2, t)$ and the dots in the fifth represent $\mathcal{P}_{dd}(\lambda_1, \lambda_2)\, k(\hat{\lambda}_1, t)\, k^*(\lambda_2, t)$. Noting $\Lambda_d \cap \Lambda_0 = \{\emptyset\}$, then the integrals containing $(\cdots)$'s are zero, whilst the first and third term cancel, leaving with the desired expression equation (4.23d). □

D

Posterior Distribution for General Nonlinear Signal Model

In section §7.7.3, the posterior distribution for the parameters $\boldsymbol{\theta}$ of a system driven by a block stationary AR model is given in the form:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x},\, \boldsymbol{\phi},\, \mathcal{I}\right) \propto \frac{p\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\phi},\, \mathcal{I}\right)}{\mathscr{J}\left(\mathbf{x}, \mathbf{s}\right)} \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T S_i \left(S_i^T S_i + \delta_i^{-2} I_{Q_i}\right)^{-1} S_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|S_i^T S_i + \delta_i^{-2} I_{Q_i}\right|^{\frac{1}{2}}} \tag{7.39a}$$

which shall herein be referred to as the 1-D case; section §11.2.1.1 produces a similar form for the posterior distribution of $\boldsymbol{\theta}$ in the 2-D system driven by a 2-D block stationary AR model. This appendix details some of the steps in deriving these posterior distributions. As noted in section §7.7.2.1, the prior distribution for the AR coefficients is given by:

$$\mathbf{b} \,|\, \sigma^2 \sim \mathcal{N}\left(\mathbf{b} \,\middle|\, \mathbf{0}_P,\, \sigma^2 \delta^2 I_P\right),\ \delta \in \mathbb{R}^+ \tag{D.1}$$

where δ is a hyperparameter. Moreover, the prior for the excitation variance is:

$$p\left(\sigma^2 \,\middle|\, \nu, \gamma\right) = \mathcal{IG}\left(\sigma^2 \,\middle|\, \frac{\nu}{2}, \frac{\gamma}{2}\right) \tag{7.30}$$

where (ν, γ) are also hyperparameters. The values for these hyperparameters are unknown and often have a prior assigned to them; these 'hyper-priors' also depend on hyper-hyperparameters. It is expected that the form of the posterior distribution is less likely to be susceptible to changes in the hyper-hyperparameters, than to changes in the hyperparameters.

A complete Bayesian hierarchical model for the 1-D BSAR case can be found in [298], which is necessary for model order selection. In the experiments presented in this dissertation, a slightly less general form of Bayesian hierarchical model is chosen since the additional terms serve to obscure the underlying principle of utilising nonstationarity in a system. As such, the hyper-parameters, $\{\nu, \gamma\}$, for the excitation variance, σ^2, are assumed to be known, and a hyperprior is placed on the AR hyperparameter, δ, such that the influence on the posterior distribution of the hyper-hyperparameters is minimal. Although, by setting $\delta^2 = 0$, it is possible to put an uninformative prior on the AR parameters, this only results in a minor simplification and is a less robust method of choosing a prior. A vague conjugate prior density is ascribed to δ^2 using an inverse-Gamma distribution:

$$\delta^2 \sim \mathcal{IG}\left(\delta^2 \,\middle|\, \alpha_{\delta^2}, \beta_{\delta^2}\right) \tag{D.2}$$

D.1 1-D Block Stationary AR Model

Proof (Proof of equation (7.39a)). Assigning the priors and hyperpriors discussed above to each data block, Bayes' rule, of equation (7.36b), is modified appropriately, and the joint distribution of all known parameters becomes:

$$p\left(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\delta} \,\middle|\, \mathbf{x}, \boldsymbol{\phi}_{-\boldsymbol{\delta}}, \mathcal{I}\right) = \frac{p\left(\boldsymbol{\theta} \,\middle|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,\middle|\, \mathcal{I}\right) \mathcal{J}(\mathbf{x}, \mathbf{s})} \prod_{i=1}^{M} \left[\frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{ -\frac{(\mathbf{s}_i + \mathbf{S}_i\mathbf{b}_i)^T(\mathbf{s}_i + \mathbf{S}_i\mathbf{b}_i)}{2\sigma_i^2} \right\} \right.$$
$$\left. \times \frac{1}{(\sqrt{2\pi}\delta_i\sigma_i)^{Q_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}} (\sigma_i^2)^{-(\frac{\nu_i}{2}+1)}}{\Gamma(\frac{\nu_i}{2})} \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}} (\delta_i^2)^{-(\alpha_{\delta_i^2}+1)}}{\Gamma(\alpha_{\delta_i^2})} \exp\left\{ -\frac{\gamma_i}{2\sigma_i^2} - \frac{\mathbf{b}_i^T\mathbf{b}_i}{2\delta_i^2\sigma_i^2} - \frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \right]$$

where $\boldsymbol{\delta} \triangleq \{\delta_i,\, i \in \mathscr{M}\}$ and $\boldsymbol{\phi}_{-\boldsymbol{\delta}}$ denotes $\boldsymbol{\phi}$ with $\boldsymbol{\delta}$ removed. This can be written as:

$$\begin{aligned} p\left(\boldsymbol{\theta},\, \boldsymbol{\psi},\, \boldsymbol{\delta} \,|\, \mathbf{x},\, \boldsymbol{\phi}_{-\boldsymbol{\delta}},\, \mathcal{I}\right) = \frac{p\left(\boldsymbol{\theta} \,|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,|\, \mathcal{I}\right)\ \mathscr{J}\left(\mathbf{x},\, \mathbf{s}\right)} \prod_{i=1}^{M} \Bigg[& \frac{1}{(\sqrt{2\pi})^{T_i+Q_i}} \frac{1}{\sigma_i^{T_i+Q_i+\nu_i+2}} \\ & \times \exp\left\{ -\frac{\gamma_i + \mathbf{s}_i^{\mathsf{T}}\,\mathbf{s}_i + 2\mathbf{b}_i^{\mathsf{T}}\,S_i^{\mathsf{T}}\,\mathbf{s}_i + \mathbf{b}_i^{\mathsf{T}}\left(S_i^{\mathsf{T}}\,S_i + \delta_i^{-2}\,I_{Q_i}\right)\,\mathbf{b}_i}{2\sigma_i^2} \right\} \\ & \times \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma\left(\frac{\nu_i}{2}\right)} \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}}}{\Gamma\left(\alpha_{\delta_i^2}\right)} \left(\delta_i^2\right)^{-\left(\frac{Q_i}{2}+\alpha_{\delta_i^2}+1\right)} \exp\left\{ -\frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \Bigg] \end{aligned}$$

The AR parameters, $\mathbf{b}$, are marginalised as in (7.38a), using the identity [127]:

$$\int_{\mathbb{R}^P} \exp\left\{ -\frac{1}{2}\left[\alpha + 2\mathbf{y}^{\mathsf{T}}\,\boldsymbol{\beta} + \mathbf{y}^{\mathsf{T}}\,\Gamma\,\mathbf{y}\right] \right\} \mathrm{d}\mathbf{y} = \frac{(2\pi)^{\frac{P}{2}}}{|\Gamma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\left[\alpha - \boldsymbol{\beta}^{\mathsf{T}}\,\Gamma^{-1}\boldsymbol{\beta}\right] \right\} \quad \text{(D.3)}$$

where $\mathbf{y} \in \mathbb{R}^P$. Therefore:

$$\begin{aligned} p\left(\boldsymbol{\theta},\, \boldsymbol{\sigma},\, \boldsymbol{\delta} \,|\, \mathbf{x},\, \boldsymbol{\phi}_{-\boldsymbol{\delta}},\, \mathcal{I}\right) = \frac{p\left(\boldsymbol{\theta} \,|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,|\, \mathcal{I}\right)\ \mathscr{J}\left(\mathbf{x},\, \mathbf{s}\right)} \prod_{i=1}^{M} \Bigg[& \frac{1}{(\sqrt{2\pi})^{T_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma\left(\frac{\nu_i}{2}\right)} \frac{1}{\left|S_i^{\mathsf{T}}\,S_i + \delta_i^{-2}\,I_{Q_i}\right|^{\frac{1}{2}}} \frac{1}{\sigma_i^{T_i+\nu_i+2}} \\ & \times \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}}}{\Gamma\left(\alpha_{\delta_i^2}\right)} \left(\delta_i^2\right)^{-\left(\frac{Q_i}{2}+\alpha_{\delta_i^2}+1\right)} \exp\left\{ -\frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \exp\left\{ -\frac{\gamma_i + \mathbf{s}_i^{\mathsf{T}}\,\mathbf{s}_i - \mathbf{s}_i^{\mathsf{T}}\,S_i\left(S_i^{\mathsf{T}}\,S_i + \delta_i^{-2}\,I_{Q_i}\right)^{-1}\,S_i^{\mathsf{T}}\,\mathbf{s}_i}{2\sigma_i^2} \right\} \Bigg] \end{aligned} \quad \text{(D.4)}$$

Finally, noting the Gamma integral, defined as [127],

$$\Gamma(\alpha) = \int_0^\infty \gamma^{\alpha-1}\, e^{-\gamma}\, \mathrm{d}\gamma \quad \text{(D.5)}$$

then, setting $\hat{\sigma}_i \propto \sigma_i^{-2}$, σ_i can be integrated out, as in equation (7.38b), to give:

$$\begin{aligned} p\left(\boldsymbol{\theta},\, \boldsymbol{\sigma},\, \boldsymbol{\delta} \,|\, \mathbf{x},\, \boldsymbol{\phi}_{-\boldsymbol{\delta}},\, \mathcal{I}\right) = \frac{p\left(\boldsymbol{\theta} \,|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,|\, \mathcal{I}\right)\ \mathscr{J}\left(\mathbf{x},\, \mathbf{s}\right)} \prod_{i=1}^{M} \Bigg[& \frac{1}{(\sqrt{2\pi})^{T_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma\left(\frac{\nu_i}{2}\right)} \frac{1}{\left|S_i^{\mathsf{T}}\,S_i + \delta_i^{-2}\,I_{Q_i}\right|^{\frac{1}{2}}} \\ & \times \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}}}{\Gamma\left(\alpha_{\delta_i^2}\right)} \left(\delta_i^2\right)^{-\left(\frac{Q_i}{2}+\alpha_{\delta_i^2}+1\right)} \exp\left\{ -\frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \frac{\Gamma(R_i)}{\left\{\gamma_i + \mathbf{s}_i^{\mathsf{T}}\,\mathbf{s}_i - \mathbf{s}_i^{\mathsf{T}}\,S_i\left(S_i^{\mathsf{T}}\,S_i + \delta_i^{-2}\,I_{Q_i}\right)^{-1}\,S_i^{\mathsf{T}}\,\mathbf{s}_i\right\}^{R_i}} \Bigg] \end{aligned} \quad \text{(D.6)}$$

where $R_i = \frac{T_i+\nu_i+1}{2}$. Note that if the prior on the hyper-parameter $\boldsymbol{\delta}$ is omitted, this expression reduces to:

$$p\left(\boldsymbol{\theta} \,\middle|\, \mathbf{x},\, \boldsymbol{\phi},\, \mathcal{I}\right) = \frac{p\left(\boldsymbol{\theta} \,\middle|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,\middle|\, \mathcal{I}\right)\ \mathscr{J}(\mathbf{x},\, \mathbf{s})} \prod_{i=1}^{M} \left[\frac{1}{(\sqrt{2\pi})^{T_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma\left(\frac{\nu_i}{2}\right)} \frac{1}{\left|S_i^T S_i + \delta_i^{-2} I_{Q_i}\right|^{\frac{1}{2}}} \right.$$
$$\left. \times \frac{\Gamma(R_i)}{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T S_i \left(S_i^T S_i + \delta_i^{-2} I_{Q_i}\right)^{-1} S_i^T \mathbf{s}_i\right\}^{R_i}} \right] \tag{D.7}$$

□

E

Implementing the Gibbs Sampler

This appendix outlines the details for the implementation of the Gibbs sampler for the investigations discussed in section §8.1. Section §8.2.1 demonstrates the underlying mechanism in the marginalisation of the BSAR parameters of the source signal for a second-order filter. It is of interest to investigate the properties of this model for higher-order channels, and this can only be achieved by simulation since an exhaustive search of the parameter space is impossible. For completeness, section §E.2 outlines how this is implemented. Details of the Gibbs sampler can be found in, for example, [283], and methods of sampling from standard distributions are discussed in, for example, [294].

E.1 Simulation of Histogram Technique

In section §8.1.1, the conditional posterior distribution is derived for the case when the distortion filter, $\mathscr{A}$, is assumed *block time-variant* in order to facilitate the histogram technique for parameter estimation. The posterior distribution, conditional

on the hyperparameters and data for the variants $\{\mathbf{c}_i,\ \sigma_i\}$, where $\mathbf{c}_i \triangleq \begin{bmatrix}\mathbf{a}_i & \mathbf{b}_i\end{bmatrix}$, is:

$$p\left(\mathbf{c}_i,\ \sigma_i^2 \,\middle|\, \mathbf{x}_i,\ \mathbf{x}_{i-1}\right) \propto \frac{1}{\left(\sqrt{2\pi}\sigma_i\right)^{\hat{R}_i}\ \delta_i^{P+Q_i}} \exp\left\{-\frac{\gamma_i + \mathbf{e}_i^T\,\mathbf{e}_i + \delta_i^{-2}\,\mathbf{c}_i^T\,\mathbf{c}_i}{2\sigma_i^2}\right\} \tag{8.6}$$

where $\hat{R}_i = T_i + P + Q_i + \nu_i + 2$, and the excitation samples in block i, $\mathbf{e}_i$, may be written as $\mathbf{e}_i = \mathbf{x}_i + \mathbf{X}_i\,\mathbf{c}_i$. Hence, $p\left(\mathbf{c}_i \,\middle|\, \mathbf{x}_i,\ \mathbf{x}_{i-1}\right)$ can be obtained by marginalising σ_i^2. However, it is difficult to sample the parameters $\mathbf{c}_i$ from the resulting distribution and, therefore, the Gibbs sampler is employed. In the investigations discussed here, an uninformative Jeffrey's prior is chosen for the excitation variance and, hence, $\nu_i = \gamma_i = 0,\ \forall i \in \mathscr{M}$. Further, to ensure that this distribution is not greatly dependent on the remaining hyperparameter δ_i, the Bayesian model is extended such that δ_i is also considered as a parameter: a hyperprior is placed on δ_i such that the influence on the posterior distribution of the hyper-hyperparameters is minimal. Although, by setting $\delta_i^2 = 0$, it is possible to put an uninformative prior on the AR parameters, this only results in a minor simplification and is a less robust method of choosing a prior. A vague conjugate prior density is ascribed to δ_i^2 using an inverse-Gamma distribution:

$$\delta_i^2 \sim \mathcal{IG}\left(\delta_i^2 \,\middle|\, \alpha_{\delta_i^2},\ \beta_{\delta_i^2}\right) \tag{E.1}$$

In order to implement the Gibbs sampler, the conditional probabilities $p\left(\mathbf{c}_i \,\middle|\, \sigma_i^2,\ \delta_i^2,\ \mathbf{x}_i,\ \mathbf{x}_{i-1}\right)$, $p\left(\sigma_i^2 \,\middle|\, \mathbf{c}_i,\ \delta_i^2,\ \mathbf{x}_i,\ \mathbf{x}_{i-1}\right)$ and $p\left(\delta_i^2 \,\middle|\, \mathbf{c}_i,\ \sigma_i^2,\ \mathbf{x}_i,\ \mathbf{x}_{i-1}\right)$ must be easily sampled from. From section §8.1.1 the likelihood function is:

$$p\left(\mathbf{x}_i \,\middle|\, \mathbf{c}_i,\ \sigma_i^2,\ \mathbf{x}_{i-1}\right) \sim \mathcal{N}\left(\mathbf{e}_i \,\middle|\, 0,\ \sigma_i^2\right) \tag{8.5}$$

E.1.1 Conditional Density for AR parameters

The required conditional density for the autoregressive parameters, $\mathbf{c}_k$, is given by $p\left(\mathbf{c}_k \,\middle|\, \sigma_k^2,\ \delta^2,\ \mathbf{x}_k,\ \mathbf{x}_{k-1}\right)$, $k \in \mathscr{M}$. Applying Bayes' rule, the posterior probability for the unknown AR parameters, $\mathbf{c}_k$, becomes:

$$p\left(\mathbf{c}_k \,\middle|\, \sigma_k^2,\ \delta_k^2,\ \mathbf{x}_k,\ \mathbf{x}_{k-1}\right) = \frac{p\left(\mathbf{x}_k \,\middle|\, \mathbf{c}_k,\ \sigma_k^2,\ \mathbf{x}_{k-1}\right)\ p\left(\mathbf{c}_k \,\middle|\, \delta_k^2\right)\ p\left(\delta_k^2 \,\middle|\, \alpha_{\delta_k^2}, \beta_{\delta_k^2}\right)}{p\left(\mathbf{x}_k,\ \mathbf{x}_{k-1}\right)} \tag{E.2}$$

where the AR parameters, $\mathbf{c}_k$, are assumed to be Gaussian distributed:

$$\mathbf{c}_k \,|\, \sigma_k^2 \sim \mathcal{N}\left(\mathbf{c}_k \,\middle|\, \mathbf{0}_{P+Q_k},\ \sigma_k^2\,\delta_k^2 \mathrm{I}_{P+Q_k}\right),\ \delta_k > 0 \tag{E.3}$$

Hence, after some slight rearrangements, equation (E.2) may be written in the form:

$$p\left(\mathbf{c}_k \,\middle|\, \sigma_k^2,\ \delta_k^2,\ \mathbf{x}_k,\ \mathbf{x}_{k-1}\right) \propto \exp\left\{-\frac{(\mathbf{c}_k - \hat{\mathbf{c}}_k)^T\,\mathbf{C}_k^{-1}\,(\mathbf{c}_k - \hat{\mathbf{c}}_k)}{2}\right\} \tag{E.4a}$$

This is a multivariate Gaussian with inverse covariance matrix given by:

$$\mathbf{C}_k^{-1} = \frac{\mathbf{X}_k^T\,\mathbf{X}_k + \delta_k^{-2}\mathrm{I}_{P+Q_k}}{\sigma_k^2} \tag{E.4b}$$

and mode $\hat{\mathbf{c}}_k$:

$$\hat{\mathbf{c}}_k = -\left(\mathbf{X}_k^T\,\mathbf{X}_k + \delta_k^{-2}\mathrm{I}_{P+Q_k}\right)^{-1}\mathbf{X}_k^T\,\mathbf{x}_k \tag{E.4c}$$

Algorithm E.1, below, describes the steps used for generating a random sample vector for the AR parameters, $\mathbf{b}_k$. Each of these steps are simple to implement [283, 294].

Algorithm E.1 Sampling Autoregressive Parameters.

1: Determine the square root, L_k, of the inverse covariance matrix $\mathbf{C}_k^{-1}$. This can be carried out using a standard Cholesky decomposition, providing the covariance matrix is positive definite [283, 294].
2: Generate a Gaussian random vector, $\mathbf{u}_k$, with each component of unit variance and with the same length as the parameter vector $\mathbf{c}_k$: *i.e.* $\mathbf{u}_k \sim \mathcal{N}\left(\mathbf{u}_k \,\middle|\, \mathbf{0},\ \mathrm{I}_{P+Q_k}\right)$.
3: Solve $L_k^T\,\mathbf{n}_k = \mathbf{u}_k$.
4: Solve $\left(\mathbf{X}_k^T\,\mathbf{X}_k + \delta_k^{-2}\mathrm{I}_{P+Q_k}\right)\hat{\mathbf{c}}_k = \mathbf{X}_k^T\,\mathbf{x}_k$.
5: Compute $\mathbf{c}_k = \hat{\mathbf{c}}_k + \mathbf{n}_k$.

E.1.2 Conditional Density for Excitation Variance

The conditional density for the excitation variance, σ_k^2, in block k is given from Bayes' rule as $p\left(\sigma_k^2 \,\middle|\, \delta_k^2,\ \mathbf{c}_k,\ \mathbf{x}_k,\ \mathbf{x}_{k-1}\right) \propto p\left(\mathbf{x}_k \,\middle|\, \mathbf{c}_k,\ \sigma_k^2,\ \mathbf{x}_{k-1}\right)\ p\left(\sigma_k^2\right)$, since there is less need to place hyperpriors on the parameters of the prior distribution. The usual prior for the standard deviation is the inverse-Gamma density:[1]

$$\sigma_k^2 \sim \mathcal{IG}\left(\sigma_k^2 \,\middle|\, \nu_k,\ \gamma_k\right) \tag{E.5}$$

[1] The inverse-Gamma is of the form given in equation (7.30) – see section §7.7.2.2 on page 179.

Hence, the conditional posterior is in the form:

$$p\left(\sigma_k^2 \,\middle|\, \delta_k^2,\, \mathbf{c}_k,\, \mathbf{x}_k,\, \mathbf{x}_{k-1}\right) \propto \frac{1}{\sigma_i^{T_k+\nu_k+2}} \exp\left\{-\frac{\mathbf{e}_k^T\,\mathbf{e}_k+\gamma_k}{2\sigma_k^2}\right\} \tag{E.6}$$

Note that since the prior is conjugate, the conditional posterior distribution is also in the form of an inverse-Gamma distribution. Although it is difficult to sample directly from this distribution, it is, however, easy to sample from the Gamma distribution, followed by application of the probability transformation rule [286]. The following steps are used to generate σ_k^2, with the required inverse-Gamma density [283]:

Algorithm E.2 Sampling an Inverse-Gamma Distribution.

1: Generate a Gamma variate with $\alpha = \frac{T_k+\alpha_k}{2}$ degrees of freedom $x_k \leftarrow x^{\alpha-1}\exp(-x)$.
2: Take the reciprocal and scale the result: *i.e.* $\sigma_k^2 = \frac{\mathbf{e}_k^T\,\mathbf{e}_k+\gamma_k}{2}\frac{1}{x_k}$.

E.1.3 Conditional Density for Variance of AR Parameter Prior

The hyper-prior for the variance of the prior on the AR parameters is also given by an inverse-Gamma distribution, $\delta_i^2 \sim \mathcal{IG}\left(\delta_i^2 \,\middle|\, \alpha_{\delta_i^2},\, \beta_{\delta_i^2}\right)$, and the resulting conditional density is:

$$p\left(\delta_k^2 \,\middle|\, \mathbf{c}_k,\, \sigma_k^2,\, \mathbf{x}_k,\, \mathbf{x}_{k-1}\right) \propto \mathcal{IG}\left(\delta_k^2 \,\middle|\, \alpha_{\delta_k^2}+\frac{P+Q_k}{2},\, \beta_{\delta_k^2}+\frac{\mathbf{c}_k^T\,\mathbf{c}_k}{2}\right) \tag{E.7}$$

which can be sampled from a Gamma distribution with inversion of the results as discussed in section §E.1.2.

E.2 Exploration of Parameter Space for Constrained Channel Model

As stated in the introduction of this appendix, it is of interest to investigate the properties of the model introduced in section §8.2.1 for higher-order channels. This can only be achieved by simulation, and this section outlines the implementation.

To explore the parameter space for $\mathbf{a}$, samples are drawn from the distribution:

$$p\left(\mathbf{a} \,\middle|\, \mathbf{x}\right) \propto p\left(\mathbf{a}\right) \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T S_i \left(S_i^T S_i + \delta_i^{-2} I_{Q_i}\right)^{-1} S_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|S_i^T S_i + \delta_i^{-2} I_{Q_i}\right|^{\frac{1}{2}}} \tag{8.7}$$

where $R_i = \frac{\nu_j + T_j + 1}{2}$, $i \in \mathscr{M}$. This distribution is difficult to sample from and, furthermore, it is desirable to sample the BSAR coefficients $\{\mathbf{b}_i,\, i \in \mathscr{M}\}$. Again, this can be efficiently implemented using the Gibbs sampler by drawing the variates $\{\mathbf{a},\, \mathbf{b}_i,\, i \in \mathscr{M}\}$ from the distribution $p\left(\mathbf{a},\, \mathbf{b}_i,\, i \in \mathscr{M} \,\middle|\, \mathbf{x}\right)$. This is only possible in practice if the full system parameter variates $\{\mathbf{a},\, \mathbf{b}_i,\, \sigma_i^2,\, i \in \mathscr{M}\}$ are drawn from the distribution $p\left(\mathbf{a},\, \mathbf{b}_i,\, \sigma_i^2,\, i \in \mathscr{M} \,\middle|\, \mathbf{x}\right)$ and a Monte Carlo method is used to marginalise the excitation variances. If a uniform prior is assigned to the excitation variance, this is analogous to ignoring the variance estimates in the histogram method. As usual, an uninformative Jeffrey's prior is chosen for the excitation variance, so $\nu_i = \gamma_i = 0,\, \forall i \in \mathscr{M}$. To ensure that the distributions are not greatly dependent on δ_i, the Bayesian model is extended such that this hyperparameter is also considered as a parameter, as discussed in section §E.1. The excitation samples in block i may either be written as:

$$\mathbf{e}_i = \mathbf{s}_i + S_i\, \mathbf{b}_i \tag{E.8a}$$

as defined in section §8.2 or, alternatively, as:

$$\mathbf{e}_i = \mathbf{y}_i + Y_i\, \mathbf{a} \tag{E.8b}$$

where $[\mathbf{y}]_{t-t_i+1} = y(t),\, t \in \mathscr{T}_i$ and $[Y]_{t-t_i,p} = y(t-p),\, t \in \mathscr{T}_i, p \in \mathscr{P}$, and $y(t) \equiv y(t, \mathbf{b}_i, \mathbf{x})$ is a function of the data, $\mathbf{x}$, and the AR parameters in block i, $\mathbf{b}_i$:

$$y(t) = x(t) + \sum_{q=1}^{Q_i} b_i(q)\, x(t-q) \tag{E.9}$$

which is the observed signal, $x(t)$, filtered by $\mathbf{b}_i$. The likelihood function is:

$$p\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta}\right) = \prod_{i=1}^{M} \mathcal{N}\left(\mathbf{e}_i \,\middle|\, 0,\, \sigma_i^2\right) \tag{E.10}$$

where $\mathbf{e}_i$ is given by either (E.8a) or (E.8b).

E.2.1 Conditional Density for BSAR Parameters

The conditional density for BSAR parameters is given by $p\left(\mathbf{b}_k \,\middle|\, \boldsymbol{\theta}_{-\mathbf{b}_k}, \mathbf{x}\right)$, $k \in \mathscr{M}$, where $\boldsymbol{\theta}_{-\mathbf{b}_k} \triangleq \{\mathbf{a}, \mathbf{b}_j, k \neq j \in \mathscr{M}, \sigma_j^2, j \in \mathscr{M}\}$. Applying Bayes' rule, the posterior probability for the unknown parameters $\mathbf{b}_k$, becomes:

$$p\left(\mathbf{b}_k \,\middle|\, \boldsymbol{\theta}_{-\mathbf{b}_k}, \mathbf{x}\right) = \frac{p\left(\mathbf{x} \,\middle|\, \boldsymbol{\theta}\right) p\left(\mathbf{b}_k\right)}{p\left(\mathbf{x}\right)} \tag{E.11a}$$

Assuming the $\mathbf{b}_k$'s are independent, $p\left(\mathbf{b}_k \,\middle|\, \boldsymbol{\theta}_{-\mathbf{b}_k}, \mathbf{x}\right) \propto p\left(\mathbf{x}_k \,\middle|\, \mathbf{x}_{k-1}, \boldsymbol{\theta}\right) p\left(\mathbf{b}_k\right)$, where the proportionality constant is a function of the other parameters, but only acts as a scaling function for the rest of the probability function. If the parameter $\mathbf{b}_k$ is assumed to be Gaussian distributed, $\mathbf{b}_k \,|\, \sigma_k^2 \sim \mathcal{N}\left(\mathbf{b}_k \,\middle|\, \mathbf{0}_{Q_k}, \sigma_k^2 \delta_k^2 \mathrm{I}_{Q_k}\right)$, $\delta_k > 0$, it follows

$$p\left(\mathbf{b}_k \,\middle|\, \boldsymbol{\theta}_{-\mathbf{b}_k}, \mathbf{x}\right) \propto \exp\left\{-\frac{\mathbf{s}_k^T \mathbf{s}_k + 2\mathbf{s}_k^T \mathbf{S}_k \mathbf{b}_k + \mathbf{b}_k^T \left(\mathbf{S}_k^T \mathbf{S}_k + \delta_k^{-2} \mathrm{I}_{Q_k}\right) \mathbf{b}_k}{2\sigma_k^2}\right\}$$

This may be written as a multivariate Gaussian $p\left(\mathbf{b}_k \,\middle|\, \boldsymbol{\theta}_{-\mathbf{b}_k}, \mathbf{x}\right) \propto \mathcal{N}\left(\mathbf{b}_k \,\middle|\, \hat{\mathbf{b}}_k, C_k\right)$ where the inverse covariance matrix is given by:

$$C_k^{-1} = \frac{\mathbf{S}_k^T \mathbf{S}_k + \delta_k^{-2} \mathrm{I}_{Q_k}}{\sigma_k^2} \tag{E.12a}$$

and mode $\hat{\mathbf{b}}_k$:

$$\hat{\mathbf{b}}_k = -\left(\mathbf{S}_k^T \mathbf{S}_k + \delta_k^{-2} \mathrm{I}_{Q_k}\right)^{-1} \mathbf{S}_k^T \mathbf{s}_k \tag{E.12b}$$

E.2.2 Conditional Density for Stationary AR Parameters

A similar analysis can be applied to derive the conditional density on the stationary parameters, $\mathbf{a}$. Defining $\boldsymbol{\theta}_{-\mathbf{a}} = \{\mathbf{b}_i, i \in \mathscr{M}, \sigma_i^2, i \in \mathscr{M}\}$ and, assuming $\mathbf{a} \,|\, \sigma_a^2 \sim \mathcal{N}\left(\mathbf{a} \,\middle|\, \mathbf{0}_P, \sigma_a^2 \mathrm{I}_P\right)$, where $\sigma_a > 0$ is a hyperparameter, then

$$p\left(\mathbf{a} \,\middle|\, \boldsymbol{\theta}_{-\mathbf{a}}, \mathbf{x}\right) \propto \prod_{i=1}^{M} \exp\left\{-\frac{\mathbf{y}_i^T \mathbf{y}_i + 2\mathbf{y}_i^T \mathbf{Y}_i \mathbf{a} + \mathbf{a}^T \left(\mathbf{S}_i^T \mathbf{S}_i + \delta^{-2}\right) \mathbf{a}}{2\sigma_i^2}\right\} \tag{E.13a}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{M} \frac{\mathbf{y}_i^T \mathbf{y}_i}{\sigma_i^2} + 2\left(\sum_{i=1}^{M} \frac{\mathbf{y}_i^T \mathbf{Y}_i}{\sigma_i^2}\right) \mathbf{a} + \mathbf{a}^T \left(\frac{1}{\sigma_a^2} + \sum_{i=1}^{M} \frac{\mathbf{Y}_i^T \mathbf{Y}_i}{\sigma_i^2}\right) \mathbf{a}\right]\right\} \tag{E.13b}$$

This may be written as a multivariate Gaussian $p\left(\mathbf{a} \,\middle|\, \boldsymbol{\theta}_{-\mathbf{a}}, \mathbf{x}\right) \propto \mathcal{N}\left(\mathbf{a} \,\middle|\, \hat{\mathbf{a}}, C_a\right)$ where the inverse covariance matrix is given by:

$$C_a^{-1} = \frac{1}{\sigma_a} + \sum_{i=1}^{M} \frac{\mathbf{Y}_i^T \mathbf{Y}_i}{\sigma_i^2} \tag{E.14a}$$

and mode $\hat{\mathbf{b}}_i$:

$$\hat{\mathbf{a}} = -\left(\frac{1}{\sigma_a} + \sum_{i=1}^{M} \frac{\mathbf{Y}_i^T \mathbf{Y}_i}{\sigma_i^2}\right)^{-1} \left(\sum_{i=1}^{M} \frac{\mathbf{Y}_i^T \mathbf{y}_i}{\sigma_i^2}\right) \tag{E.14b}$$

E.2.3 Conditional Density for Time-Varying Excitation Variance

Finally, the conditional probabilities $p\left(\sigma_k^2 \,\middle|\, \boldsymbol{\theta}_{-\sigma_k}, \mathbf{x}\right)$, where $\boldsymbol{\theta}_{-\sigma_k} = \{\mathbf{a}, \mathbf{b}_i, i \in \mathscr{M}, \sigma_j^2, k \neq j \in \mathscr{M}\}$, is required. The usual prior for the standard deviation is the inverse-Gamma density and, hence, using Bayes' rule, setting $\alpha_k \rightarrow \frac{\alpha_k}{2}$ and $\beta_k \rightarrow \frac{\beta_k}{2}$:

$$p\left(\sigma_k^2 \,\middle|\, \boldsymbol{\theta}_{-\sigma_k} \mathbf{x}\right) \propto \frac{1}{\sigma_k^{T_k+\alpha_k-2}} \exp\left\{-\frac{\mathbf{e}_k^T \mathbf{e}_k + \beta_k}{2\sigma_k^2}\right\} \tag{E.15}$$

where $\mathbf{e}_k$ is given by (E.8). It may be shown, by differentiation, that the mode is:

$$\hat{\sigma}_k = \sqrt{\frac{\mathbf{e}_k^T \mathbf{e}_k}{N}} \tag{E.16}$$

It is difficult to sample directly the density in equation (E.15). However, samples can be generated form a Gamma distribution and transformed as discussed in section §E.1.2.

Bibliography

[1] ABDRABBO, N. A. and M. B. PRIESTLEY, "On the prediction of non-stationary processes," *Journal of the Royal Statistical Society – Series B*, vol. 29, no. 3, pp. 570–585, 1967.

[2] ABDRABBO, N. A. and M. B. PRIESTLEY, "Filtering non-stationary processes," *Journal of the Royal Statistical Society – Series B*, vol. 31, no. 1, pp. 150–159, 1969.

[3] ABED-MERAIM, K., W. QIU, and Y. HUA, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.

[4] AHMED, A., *Signal Separation*, Ph. D. Thesis, University of Cambridge, UK, Sept. 2000.

[5] AHMED, A., P. J. W. RAYNER, and S. J. GODSILL, "Considering non-stationarity for blind signal separation," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 111–114, Mohonk Mountain House, New York, Oct. 1999.

[6] AL-CHALABI, M., "When least-squares square least," *Geophysical Prospecting*, vol. 40, no. 3, pp. 359–378, Apr. 1992.

[7] AL-SHOSHAN, A. I. and L. F. CHAPARRO, "Identification of non-minimum phase systems using evolutionary spectral theory," *Signal Processing*, vol. 55, no. 1, pp. 79–92, Nov. 1996.

[8] ALLEN, J. B. and D. A. BERKLEY, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943 – 950, Apr. 1979.

[9] ALLEN, J. B., D. A. BERKLEY, and J. BLAUERT, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *Journal of the Acoustical Society of America*, vol. 62, no. 4, pp. 912–915, Oct. 1977.

[10] ALMEIDA, L. B. and J. M. TRIBOLET, "Nonstationary spectral modeling of voiced speech," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-31, no. 3, pp. 664–678, June 1983.

[11] AMIN, M. G., "Time-varying spectrum estimation for a general class of nonstationary processes," *Proc. IEEE*, vol. 74, no. 12, pp. 1800–1802, Dec. 1986.

[12] ANDRIEU, C., E. BARAT, and A. DOUCET, "Bayesian deconvolution of noisy filtered point processes," Technical Report CUED/F-INFENG/TR. 352, Department of Engineering, University of Cambridge, UK, July 1999, accepted for publication in IEEE Transactions on Signal Processing, 2000.

[13] ANDRIEU, C., P. M. DJURIĆ, and A. DOUCET, "Model selection by MCMC computation," *Signal Processing*, 2000, to appear.

[14] ANDRIEU, C. and A. DOUCET, "A Bayesian approach to harmonic retrieval with clipped data," *Signal Processing*, vol. 74, no. 3, pp. 239–252, May 1999.

[15] ANDRIEU, C. and A. DOUCET, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2667–2676, 1999.

[16] ARAKAWA, K., D. H. FENDER, H. HARASHIMA, H. MIYAKAWA, and Y. SAITOH, "Separation of a nonstationary component from the EEG by a nonlinear digital filter," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, no. 7, pp. 724–726, July 1986.

[17] ASELTINE, J. A., "A transform method for linear time-varying systems," *Journal Applied Physics*, vol. 25, no. 6, pp. 761–764, June 1954.

[18] ATAL, B. S. and S. L. HANAUER, "Speech analysis and synthesis by linear prediction of the speech wave," *Journal of the Acoustical Society of America*, vol. 50, no. 2, pp. 637–655, Aug. 1971.

[19] ATLAS, L. and P. DUHAMEL, "Recent developments in the core of digital signal processing," *IEEE Signal Processing Magazine*, vol. 16, no. 1, pp. 16–31, Jan. 1999.

[20] AUDRE-OBRECHT, R., "A new statistical approach for the automatic segmentation of continuous speech signals," *IEEE Transactions on Speech, and Signal Processing*, vol. 36, no. 1, pp. 29 – 40, Jan. 1988.

[21] BALL, J. A., I. GOHBERG, and M. A. KAASHOEK, "A frequency response function for linear, time-varying systems," *Mathematics of Control Signals and Systems*, vol. 8, no. 4, pp. 334–351, 1995.

[22] BARAM, Y., "Identifying nonstationary measurement noise in linear systems," *IEEE Transactions on Information Theory*, vol. IT-28, no. 1, pp. 122–123, Jan. 1982.

[23] BASSEVILLE, M. and A. BENVENISTE, "Design and comparative study of some sequential jump detection algorithms for digital signals," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-31, no. 3, pp. 521 – 535, June 1983.

[24] BAZARAA, M. S., J. J. JARVIS, and H. D. SHERALI, *Linear Programming and Network Flows*, John Wiley & Sons, Inc., 1990.

[25] BEAUGEANT, C., V. TURBIN, P. SCALART, and A. GILLOIRE, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing*, vol. 64, no. 1, pp. 33–47, Jan. 1998.

[26] BEES, D., M. BLOSTEIN, and P. KABAL, "Reverberant speech enhancement using cepstral processing," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 977–980, Toronto, Ont., Canada, May 1991.

[27] BELAL, A. A. and B. A. SHENOI, "Passive realization of a class of time-varying networks having a compatible transform," *IEEE Transactions on Circuits and Systems*, vol. CAS-27, no. 7, pp. 644–646, July 1980.

[28] BELAL, A. A. and B. A. SHENOI, "Properties of a compatible transform for a class of time-varying digital filters." *IEEE Transactions on Circuits and Systems*, vol. CAS-28, no. 10, pp. 1020–1024, Oct. 1981.

[29] BELAL, A. A. and B. A. SHENOI, "Frequency scaling of a linear time-invariant network by a time-varying function," *IEEE Transactions on Circuits and Systems*, vol. CAS-29, no. 1, pp. 57–58, Jan. 1982.

[30] BELICZYNSKI, B., I. KALE, and G. D. CAIN, "Approximation of FIR by IIR digital filters: an algorithm based on balanced model reduction," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 532–542, Mar. 1992.

[31] BELLO, P. A., "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communication Systems*, vol. 11, pp. 360–393, Dec. 1963.

[32] BELLO, P. A., "Time-frequency duality," *IRE Transactions on Information Theory*, vol. IT-10, pp. 18–33, Jan. 1964.

[33] BENDAT, J., "A general theory of linear prediction and filtering," *Journal of the Society for Industrial and Applied Mathematics*, vol. 4, no. 3, pp. 131–151, Sept. 1956.

[34] BERKHOUT, A. J., D. DE VRIES, and M. M. BOONE, "A new method to acquire impulse responses in concert halls," *Journal of the Acoustical Society of America*, vol. 68, no. 1, pp. 179–183, July 1980.

[35] BERNARDO, J. M. and A. F. M. SMITH, *Bayesian Theory*, John Wiley & Sons, Inc., 1994.

[36] BESSON, O. and F. CASTANIE, "On estimating the frequency of a sinusoid in autoregressive multiplicative noise," *Signal Processing*, vol. 30, no. 1, pp. 65–83, Jan. 1993.

[37] BESSON, O., M. GHOGHO, and A. SWAMI, "On estimating random amplitude chirp signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1561–1564, Phoenix, Arizona, Mar. 1999.

[38] BIEMOND, J., R. L. LAGENDIJK, and R. M. MERSEREAU, "Iterative methods for image deblurring," *Proc. IEEE*, vol. 78, no. 5, pp. 856–883, May 1990.

[39] BIEMOND, J., F. G. VAN DER PUTTEN, and J. W. WOODS, "Identification and restoration of images with symmetric noncausal blurs," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 4, pp. 385–393, Apr. 1988.

[40] BISTRITZ, Y., "Zero location with respect to the unit circle of discrete-time linear system polynomials," *Proc. IEEE*, vol. 72, no. 9, pp. 1131–1142, Sept. 1984.

[41] BISTRITZ, Y., "Comments on "On zero location with respect to the unit circle of discrete-time linear system polynomials"," *Proc. IEEE*, vol. 74, no. 12, pp. 1802–1803, Dec. 1986.

[42] BLOOM, P. J. and G. D. CAIN, "Evaluation of two-input speech dereverberation techniques," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 164–167, Paris, France, May 1982.

[43] BODE, H. W. and C. E. SHANNON, "A simplified derivation of linear least square smoothing and prediction theory," *Proc. IRE*, vol. 38, pp. 417–425, Apr. 1950.

[44] BOLLERSLEV, T., "Generalised autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 31, pp. 307 – 327, 1986.

[45] BOLLERSLEV, T., R. Y. CHOU, and K. F. KRONER, "ARCH modelling in finance: A review of the theory of empirical evidence," *Journal of Econometrics*, vol. 52, pp. 5 – 59, 1992.

[46] BOOTON, JR., R. C., "An optimisation theory for time-varying linear systems with nonstationary statistical inputs," *Proc. IRE*, vol. 40, pp. 977–981, Aug. 1952.

[47] BOUDAOUD, M. and L. F. CHAPARRO, "Composite modeling of nonstationary signals," *Journal of the Franklin Institute*, vol. 324, no. 1, pp. 113–124, 1987.

[48] BOUDAOUD, M. and L. F. CHAPARRO, "Nonstationary composite modeling of images," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 1, pp. 112–117, Jan. - Feb. 1989.

[49] BOX, G. E. P., G. M. JENKINS, and G. C. REINSEL, *Time Series Analysis: Forecasting and Control*, Holden-Day, third edition, 1994.

[50] BOYLES, R. A. and W. A. GARDNER, "Cycloergodic properties of discrete-parameter nonstationary stochastic processes," *IEEE Transactions on Information Theory*, vol. IT-29, no. 1, pp. 105–114, Jan. 1983.

[51] BRADLEY, J. S., "Predictors of speech intelligibility in rooms," *Journal of the Acoustical Society of America*, vol. 80, no. 3, pp. 837–845, Sept. 1986.

[52] BRADLEY, J. S., R. D. REICH, and S. G. NORCROSS, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1820–1828, Oct. 1999.

[53] BROERSEN, P. M. T. and H. E. WENSINK, "On the penalty factor for autoregressive order selection in finite samples," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 748–752, Mar. 1996.

[54] BRONKHORST, A. W., "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, no. 1, pp. 117–128, Jan. - Feb. 2000.

[55] BROWN, R. G. and P. Y. C. HWANG, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, Inc., Chichester, UK, second edition, 1992.

[56] CANAGARAJAH, C. N., *Digital Signal Processing Techniques for Speech Enhancement in Hearing Aids*, Ph. D. Thesis, University of Cambridge, UK, May 1993.

[57] CAPPÉ, O., A. DOUCET, M. LAVIELLE, and E. MOULINES, "Simulation-based methods for blind maximum-likelihood filter identification," *Signal Processing*, vol. 73, no. 1-2, pp. 3–25, Feb. 1999.

[58] CASACUBERTA, F. and E. VIDAL, "A nonstationary model for the analysis of transient speech signals," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-35, no. 2, pp. 226–228, Feb. 1987.

[59] CHAMPAGNE, B., A. LOBO, and P. KABAL, "Efficient methods for simulating a moving talker in a rectangular room," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New York, 1991.

[60] CHAMPENEY, D. C., *A handbook of Fourier Theorems*, Cambridge University Press, Cambridge, 1990.

[61] CHAN, D. C. B., *Blind Signal Separation*, Ph. D. Thesis, University of Cambridge, UK, Jan. 1997.

[62] CHANG, P., C. G. LIN, and B. YEH, "Inverse filtering of a loudspeaker and room acoustics using time-delay neural networks," *Journal of the Acoustical Society of America*, vol. 95, no. 6, pp. 3400–3408, June 1994.

[63] CHARBONNIER, R., M. BARLAUD, G. ALENGRIN, and J. MENEZ, "Results on AR-modelling of nonstationary signals," *Signal Processing*, vol. 12, no. 2, pp. 143–151, Mar. 1987.

[64] CHAZAN, D., Y. MEDAN, and U. SHVADRON, "Noise cancellation for hearing aids," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 977–980, Toyko, Japan, 1986.

[65] CHAZAN, D., Y. MEDAN, and U. SHVADRON, "Noise cancellation for hearing aids," *IEEE Transactions on Speech, and Signal Processing*, vol. 36, no. 11, pp. 1697–1705, Nov. 1988.

[66] CHELLAPPA, R. and S. CHATTERJEE, "Classification of textures using gaussian markov random fields," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-33, no. 4, pp. 959–963, Aug. 1985.

[67] CHELLAPPA, R. and R. L. KASHYAP, "Digital image restoration using spatial interaction models," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-30, no. 3, pp. 461–472, June 1982.

[68] CHELLAPPA, R. and R. L. KASHYAP, "Texture synthesis using 2-D noncausal autoregressive models," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-33, no. 1, pp. 194–203, Feb. 1985.

[69] CHEN, A., S. VASEGHI, and P. MCCOURT, "State based sub-band LP Wiener filters for speech enhancement in car environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 213–216, Istanbul, Turkey, June 2000.

[70] CHEN, R. and T. H. LI, "Blind restoration of linearly degraded discrete signals by Gibbs sampling," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2410–2413, Oct. 1995.

[71] CHERRY, E. C., "Some experiments on the recognition of speech, with one and with two ears," *Journal of the Acoustical Society of America*, vol. 25, pp. 975–979, 1953.

[72] CHERRY, E. C. and W. K. TAYLOR, "Some further experiments upon recognition of speech with one and with two ears," *Journal of the Acoustical Society of America*, vol. 26, pp. 554–559, 1954.

[73] CLAASEN, T. A. C. M. and W. F. G. MECKLENBRÄUKER, "On stationary linear time-varying systems," *IEEE Transactions on Circuits and Systems*, vol. CAS-29, no. 3, pp. 169–184, Mar. 1982.

[74] CLARKSON, P. M., J. N. MOURJOPOULOS, and J. K. HAMMOND, "Spectral, phase and transient equalization of audio systems," *Journal of the Audio Engineering Society*, vol. 33, no. 3, pp. 127–132, Mar. 1985.

[75] COHEN, L., "Time-frequency distributions – a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941 – 981, July 1989.

[76] COHEN, L., "A general approach for obtaining joint representations in signal analysis. I. Characteristic function operator method," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1080–1090, May 1996.

[77] COHEN, L., "A general approach for obtaining joint representations in signal analysis. II. General class, mean and local values, and bandwidth," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1091–1098, May 1996.

[78] COHEN, L., "Generalization of the Wiener-Khinchin theorem," *IEEE Signal Processing Letters*, vol. 5, no. 11, pp. 292–294, Nov. 1998.

[79] COHEN, L., "Time-frequency analysis," in *IEEE Signal Processing Magazine* [19], pp. 22–28.

[80] COLE, D., M. MOODY, and S. SRIDHARAN, "Position-independent enhancement of reverberant speech," *Journal of the Audio Engineering Society*, vol. 45, no. 3, pp. 142–147, Mar. 1997.

[81] CRAMÉR, H., *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[82] CRAMÉR, H., "On some classes of nonstationary stochastic processes," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 57–78, University of California Press, Berkeley, California, 1961.

[83] DALIANIS, S. A. and J. K. HAMMOND, "Time-frequency spectra for frequency-modulated processes," *Mechanical Systems and Signal Processing*, vol. 11, no. 4, pp. 621 – 635, July 1997.

[84] D'ANGELO, H., *Linear Time-Varying Systems: Analysis and Synthesis*, Allyn and Bacon, Boston, 1970.

[85] DARLINGTON, S., "Nonstationary smoothing and prediction using network theory concepts," *IRE Transactions on Circuit Theory (special suppl.)*, vol. CT-6, pp. 1–11, May 1959.

[86] DAUBECHIES, I., "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, Sept. 1990.

[87] DAVIS, R. C., "On the theory of prediction of nonstationary processes," *Journal Applied Physics*, vol. 23, no. 9, pp. 1047–1053, Sept. 1952.

[88] DE CHEVEIGNE, A., "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," *Journal of the Acoustical Society of America*, vol. 93, no. 6, pp. 3271–3290, June 1993.

[89] DETKA, C. S. and A. EL-JAROUDI, "The transitory evolutionary spectrum," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. IV, pp. 289–292, Adelaide, SA, Australia, 1994.

[90] DETKA, C. S. and A. EL-JAROUDI, "Variable frequency evolutionary spectrum," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1553–1556, Detroit, MI, USA, 1995.

[91] DETKA, C. S. and A. EL-JAROUDI, "The generalized evolutionary spectrum," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2877–2881, Nov. 1996.

[92] DICKIE, J. R. and A. K. NANDI, "A comparative study of AR order selection methods," *Signal Processing*, vol. 40, no. 2-3, pp. 239–255, Nov. 1994.

[93] DJURIĆ, P. M. and S. M. KAY, "Parameter estimation of chirp signals," *IEEE Transactions on Speech, and Signal Processing*, vol. 38, no. 12, pp. 2118–2126, Dec. 1990.

[94] DJURIĆ, P. M. and S. M. KAY, "Order selection of autoregressive models," *IEEE Transactions on Signal Processing*, vol. 40, no. 11, pp. 2829–2833, Nov. 1992.

[95] DJURIĆ, P. M., S. M. KAY, and G. F. BOUDREAUX-BARTELS, "Segmentation of nonstationary signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 161–164, San Francisco, CA, USA, Mar. 1992.

[96] DOHERTY, B., S. VASEGHI, and P. MCCOURT, "Full covariance modelling and adaptation in sub-bands," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II969–II962, Istanbul, Turkey, June 2000.

[97] DONATI, F., "Finite-time averaged power spectra," *IEEE Transactions on Information Theory*, vol. 17, no. 1, pp. 7–16, Jan. 1971.

[98] DOROSLOVAČKI, M. I. and H. FAN, "Wavelet-based linear system modeling and adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1156–1167, May 1996.

[99] DUQUESNOY, A. J., "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 739–743, Sept. 1983.

[100] DUQUESNOY, A. J. and R. PLOMP, "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," *Journal of the Acoustical Society of America*, vol. 68, no. 2, pp. 537–544, Aug. 1980.

[101] EMRESOY, M. K. and A. EL-JAROUDI, "Evolutionary Burg spectral estimation," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 173–175, June 1997.

[102] ENGLE, R., "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, July 1982.

[103] FARKASH, S. and S. RAZ, "The legality problem of linear systems in Gabor time-frequency space," *Signal Processing*, vol. 34, no. 3, pp. 283–295, Dec. 1993.

[104] FARKASH, S. and S. RAZ, "Linear systems in Gabor time-frequency space," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 611–617, Mar. 1994.

[105] FONTANA, F., L. GIBIN, D. ROCCHESSO, and O. BALLAN, "Common pole equalization of small rooms using a two-step real-time digital equalizer," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 195–198, Mohonk Mountain House, New York, Oct. 1999.

[106] FOSTER, S., "Impulse response measurement using Golay codes," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 929–932, Tokyo, Japan, Apr. 1986.

[107] FRANASZEK, P. A. and B. LIU, "On a class of linear time-varying filters," *IEEE Transactions on Information Theory*, vol. IT-13, no. 3, pp. 477–484, July 1967.

[108] GABOR, D., "Theory of communication," *Journal of the Institution of Electrical Engineers*, vol. 93, pp. 429–457, 1946.

[109] GAMBOA, F. and E. GASSIAT, "Blind deconvolution of discrete linear systems," *The Annals of Statistics*, vol. 24, no. 5, pp. 1964–1981, 1996.

[110] GARDNER, W. A., "Correlation estimation and time-series modeling for nonstationary processes," *Signal Processing*, vol. 15, no. 1, pp. 31–41, July 1988.

[111] GARDNER, W. G., "Reverberation algorithms," in Kahrs and Brandenburg [172], pp. 85–132.

[112] GEEST, E. D. and R. GARCEA, "Simulation of room transmission functions using a triangular beam tracing computer model," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 253–256, Mohonk Mountain House, New York, Oct. 1995.

[113] GEORGE, E. B. and M. J. T. SMITH, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 389–406, Sept. 1997.

[114] GERSHO, A., "Characterization of time-varying linear systems," *Proc. IEEE*, vol. 51, p. 238, Jan. 1963.

[115] GHOGHO, M. and B. GAREL, "Maximum likelihood estimation of amplitude-modulated time series," *Signal Processing*, vol. 75, no. 2, pp. 99–116, June 1999.

[116] GILLOIRE, A., P. SCALART, C. LAMBLIN, C. MOKBEL, and S. PROUST, "Innovative speech processing for mobile terminals: an annotated bibliography," *Signal Processing*, vol. 80, no. 7, pp. 1149–1166, July 2000.

[117] GODSILL, S. J., *The Restoration of Degraded Audio Signals*, Ph. D. Thesis, University of Cambridge, UK, Dec. 1993.

[118] GODSILL, S. J., "Bayesian enhancement of speech and audio signals which can be modelled as ARMA processes," *International Statistical Review*, vol. 65, no. 1, pp. 1–21, 1997.

[119] GODSILL, S. J. and C. ANDRIEU, "Bayesian separation and recovery of convolutively mixed autoregressive sources," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1733–1736, Phoenix, Arizona, Mar. 1999.

[120] GODSILL, S. J. and P. J. W. RAYNER, *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag, Germany, 1998.

[121] GODSILL, S. J. and P. J. W. RAYNER, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 352–372, July 1998.

[122] GODSILL, S. J., P. J. W. RAYNER, and O. CAPPÉ, "Digital audio restoration," in Kahrs and Brandenburg [172], pp. 133–194.

[123] GODSILL, S. J. and C. H. TAN, "Removal of low frequency transient noise from old recordings using model-based signal separation techniques," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New York, Oct. 1997.

[124] GOODMAN, D. M. and M. O. EKSTROM, "Multidimensional spectral factorization and unilateral autoregressive models," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 2, pp. 258–262, Apr. 1980.

[125] GOODWIN, M. M., *Adaptive Signal Models: Theory, Algorithms and Audio Applications*, Kluwer Academic Publishers, Boston, 1998.

[126] GRACE, O. D., "Instantaneous power spectra," *Journal of the Acoustical Society of America*, vol. 69, no. 1, pp. 191–198, Jan. 1981.

[127] GRADSHTEYN, I. S. and I. M. RYZHIK, *Table of Integrals, Series, and Products*, Academic Press, Inc., fifth edition, 1994, alan Jeffrey, editor.

[128] GRANT, K. W. and P.-F. SEITZ, "The use of visible speech cues for improving auditory detection of spoken sentences," *Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1197–1208, Sept. 2000.

[129] GREEN, P. J., "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[130] GRENIER, Y., "Time-dependent ARMA modelling of nonstationary signals," *IEEE Transactions on Speech, and Signal Processing*, vol. 31, no. 4, pp. 899–911, Aug. 1983.

[131] GRENIER, Y., "Non-stationary signal modelling with application to bat echolocation calls," *Acustica*, vol. 61, no. 3, pp. 155–165, Sept. 1986.

[132] GUDVANGEN, S. and S. J. FLOCKTON, "Comparison of pole-zero and all-zero modelling of acoustic transfer functions," *Electronic Letters*, vol. 28, no. 21, pp. 1976–1978, Oct. 1992.

[133] GUDVANGEN, S. and S. J. FLOCKTON, "Modelling of acoustic transfer functions for echo cancellers," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 142, no. 1, pp. 47–51, Feb. 1995.

[134] GUSTAFSSON, F., "The marginalized likelihood ratio test for detecting abrupt changes," *IEEE Transactions on Automatic Control*, vol. 41, no. 1, pp. 66–78, Jan. 1996.

[135] GUSTAFSSON, F. and H. HJALMARSSON, "Twenty-one ML estimators for model selection," *Automatica*, vol. 31, no. 10, pp. 1377–1392, Oct. 1995.

[136] GUSTAFSSONA, S., R. MARTINA, and P. VARYA, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, no. 1, pp. 21–32, Jan. 1998.

[137] HA, P. and S. ANN, "Robust time-varying parametric modelling of voiced speech," *Signal Processing*, vol. 42, no. 3, pp. 311–317, Mar. 1995.

[138] HALL, M. G., A. V. OPPENHEIM, and A. S. WILLSKY, "Time-varying parametric modeling of speech," *Signal Processing*, vol. 5, no. 3, pp. 267–285, May 1983.

[139] HAMMOND, J. K., "Evolutionary spectra in random vibrations (with discussion)," *Journal of the Royal Statistical Society – Series B*, vol. 35, pp. 167 – 188, 1973.

[140] HAMMOND, J. K. and J. N. MOURJOPOULOS, "Cepstral methods applied to the analysis of room impulse response," in *Proceedings of The Institute of Acoustics*, pp. 51–54, Windermere, 1980.

[141] HAMMOND, J. K. and P. R. WHITE, "The analysis of non-stationary signals using time-frequency methods," *Journal of Sound and Vibration*, vol. 190, no. 3, pp. 419–447, Feb. 1996.

[142] HANEDA, Y., Y. KANEDA, and N. KITAWAKI, "Common-Acoustical-Pole and Residue model and its application to spatial interpolation and extropolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 709–717, Nov. 1999.

[143] HANEDA, Y., S. MAKINO, and Y. KANEDA, "Modeling of a room transfer function using common acoustical poles," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–213 –II–216, San Francisco, CA, USA, Mar. 1992.

[144] HANEDA, Y., S. MAKINO, and Y. KANEDA, "Common acoustical pole and zero modelling of room transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 320–328, Apr. 1994.

[145] HANEDA, Y., S. MAKINO, and Y. KANEDA, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 4, pp. 325–333, July 1997.

[146] HANEDA, Y., S. MAKINO, Y. KANEDA, and N. KITAWAKI, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 188–196, Mar. 1999.

[147] HÄNSLER, E., "The hands-free telephone problem – An annotated bibliography," *Signal Processing*, vol. 27, no. 3, pp. 259–271, June 1992.

[148] HÄNSLER, E., "Acoustic echo and noise control – It remains a challenge (Editorial)," *Signal Processing*, vol. 64, no. 1, pp. 1–2, Jan. 1998.

[149] HANSON, B. A. and D. Y. WONG, "The harmonic magnitude suppression (HMS) technique for intelligibility enhancement in the presence of interfering speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 18A.5.1–18A.5.4, San Diego, CA, USA, Mar. 1984.

[150] HARMUTH, H. F., "A generalized concept of frequency and some applications," *IEEE Transactions on Information Theory*, vol. IT-14, no. 3, pp. 375–382, 1968.

[151] HARVEY, A. C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1989.

[152] HAYKIN, S., editor, *Blind Deconvolution*, Prentice Hall, Inc., Englewood Cliff, NJ, 1994.

[153] HAZAS, M., *Processing of Non-Stationary Audio Signals*, M. Phil. Thesis, Department of Engineering, University of Cambridge, UK, Aug. 1999.

[154] HERMANSEN, K., F. K. FINK, and U. HARTMANN, "Hearing aids for profoundly deaf people based on a new parametric concept," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 89–92, Mohonk Mountain House, New York, Oct. 1993.

[155] HIROBAYASHI, S., H. NOMURA, T. KOIKE, and M. TOHYAMA, "Speech waveform recovery from a reverberant speech signal using inverse filtering of the power envelope transfer function," *Electronics and CommunicatiAutomatica,ons in Japan, Part 3 – Fundamental Electronic Science*, vol. 83, no. 6, pp. 77–85, 2000.

[156] HLAWATSCH, F., G. MATZ, H. KIRCHAUER, and W. KOZEK, "Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1417–1432, May 2000.

[157] HOFFMAN, M. W., T. D. TRINE, K. M. BUCKLEY, and D. J. VAN TASELL, "Robust adaptive microphone array processing for hearing aids: realistic speech enhancement," *Journal of the Acoustical Society of America*, vol. 96, no. 2, pp. 759–770, Aug. 1994.

[158] HOMER, J., R. R. BITMEAD, and I. MAREELS, "Quantifying the effects of dimension on the convergence rate of the LMS adaptive FIR estimator," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2611–2615, Oct. 1998.

[159] HOPFIELD, J. J., "Olfactory computation and object perception," *Proc. National academy of Sciences (USA)*, vol. 88, no. 15, pp. 6462–6466, Aug. 1991.

[160] HOPGOOD, J. R., "Blind deconvolution with application for reverberation cancellation in hearing aids," Final-year undergraduate project in Group E, Department of Engineering, University of Cambridge, UK, May 1997.

[161] HOPGOOD, J. R. and P. J. W. RAYNER, "Bayesian single channel blind deconvolution using parametric signal and channel models," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 151 – 154, Mohonk Mountain House, New York, Oct. 1999.

[162] HOPGOOD, J. R. and P. J. W. RAYNER, "Single channel signal separation using linear time-varying filters: Separability of non-stationary stochastic signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1449–1452, Phoenix, Arizona, Mar. 1999.

[163] HOPGOOD, J. R. and P. J. W. RAYNER, "Blind single channel deconvolution using nonstationary signal processing," *IEEE Transactions on Speech and Audio Processing*, 2001, submitted Nov. 2000.

[164] HOPGOOD, J. R. and P. J. W. RAYNER, "Single channel nonstationary signal separation using linear time-varying filters," *IEEE Transactions on Signal Processing*, 2001, submitted Nov. 2000.

[165] HUNT, B. R. and T. M. CANNON, "Nonstationary assumptions for Gaussian models of images," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, no. 12, pp. 876–882, Dec. 1976.

[166] JAIN, A. K., "Advances in mathematical models for image processing," *Proc. IEEE*, vol. 69, no. 5, pp. 502–528, May 1981.

[167] JAIN, A. K., *Fundamentals of Digital Image Processing*, Prentice Hall, Inc., Englewood Cliff, NJ, 1989.

[168] JAN, E. E. and J. FLANAGAN, "Image model for computer simulation of sound wave behaviour in an enclosure," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 119–122, Mohonk Mountain House, New York, Oct. 1995.

[169] JOHANSEN, L. G. and P. RUBAK, "The excess phase in loudspeaker/room transfer functions: Can it be ignored in equalization tasks?" in *Journal of the Audio Engineering Society (Abstracts)*, May 1996, preprint 4181. Presented at the 100th Convention of the Audio Engineering Society, Copenhagen, Denmark.

[170] JOHANSEN, L. G. and P. RUBAK, "Investigating speech quality by homomorphic deconvolution," in *1st European Conference on Signal prediction and Analysis*, pp. 327–330, Prague, June 1997.

[171] KAELIN, A., A. LINDGREN, and S. WYRSCH, "A digital frequency-domain implementation of a very high gain hearing aid with compensation for recruitment of loudness and acoustic echo cancellation," *Signal Processing*, vol. 64, no. 1, pp. 71–85, Jan. 1998.

[172] KAHRS, M. and K. BRANDENBURG, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer Academic Publishers, Boston, 1998.

[173] KALE, I., J. P. MACKENZIE, and T. I. LAAKSO, "Motor car acoustic response modelling and order reduction via balanced model truncation," *Electronic Letters*, vol. 32, no. 11, pp. 965–966, May 1996.

[174] KALMAN, R. E., "A new approach to linear filtering and prediction problems," *Transactions of the ASME – Journal of Basic Engineering*, vol. 83, pp. 33–45, Mar. 1960.

[175] KAMEN, E. W., "The poles and zeros of a linear time-varying system," *Linear Algebra and its Applications*, vol. 98, pp. 263–289, Jan. 1988.

[176] KATES, J. M., "Feedback cancellation in hearing aids: Results from computer simulation," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 553–562, Mar. 1991.

[177] KATES, J. M., "Superdirective arrays for hearing-aids," *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1930–1933, Oct. 1993.

[178] KATES, J. M., "Classification of background noises for hearing-aid applications," *Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 461–470, Jan. 1995.

[179] KATES, J. M., "Relating change in signal-to-noise ratio to array gain for microphone arrays used in rooms," *Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2388–2390, Apr. 1997.

[180] KATES, J. M., "Signal processing for hearing aids," in Kahrs and Brandenburg [172], pp. 235–278.

[181] KATES, J. M., "Constrained adaptation for feedback cancellation in hearing aids," *Journal of the Acoustical Society of America*, vol. 106, no. 2, pp. 1010–1019, Aug. 1999.

[182] KATES, J. M. and M. R. WEISS, "A comparison of hearing-aid array-processing techniques," *Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 3138–3148, May 1996.

[183] KATSAGGELOS, A. K. and K. T. LAY, "Maximum likelihood blur identification and image restoration using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 729–733, Mar. 1991.

[184] KAVEH, M. A. and G. R. COOPER, "An empirical investigation of the properties of the autoregressive spectral estimator," *IEEE Transactions on Information Theory*, vol. IT-22, no. 3, pp. 313–323, May 1976.

[185] KAY, S. M., *Modern Spectral Estimation*, Prentice Hall, Inc., 1988.

[186] KAY, S. M. and J. MAKHOUL, "On the statistics of the estimated reflection coefficients of an autoregressive process," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-31, no. 6, pp. 1447–1455, Dec. 1983.

[187] KAYHAN, A. S., A. EL-JAROUDI, and L. F. CHAPARRO, "Evolutionary periodogram for nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1527–1536, June 1994.

[188] KEMPÉ DE FÉRIET, J. and F. N. FRENKIEL, "Correlations and spectra for nonstationary random functions," *Mathematics of Computation*, vol. 16, no. 77, pp. 1–21, 1962.

[189] KHALIL, M. and J. DUCHÊNE, "Detection and classification of multiple events in piecewise stationary signals: Comparison between autoregressive and multiscale approaches," *Signal Processing*, vol. 75, no. 3, pp. 239–251, June 1999.

[190] KHAN, H. A. and L. F. CHAPARRO, "Formulation and implementation of the nonstationary evolutionary Wiener filtering," *Signal Processing*, vol. 76, no. 3, pp. 253–267, Aug. 1999.

[191] KINSLER, L. E., A. R. FREY, A. B. COPPENS, and J. V. SANDERS, *Fundamentals of Acoustics*, John Wiley & Sons, Inc., Chichester, UK, fourth edition, 2000.

[192] KIRCHAUER, H., F. HLAWATSCH, and W. KOZEK, "Time-frequency formulation and design of nonstationary Wiener filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1549–1552, Detroit, MI, USA, May 1995.

[193] KITAGAWA, G. and W. GERSCH, *Smoothness Priors, Analysis of Time Series*, Springer-Verlag, 1996.

[194] KLAFTER, R. D., "A compatible transform for a class of linear time-varying discrete systems," *IEEE Transactions on Circuit Theory (Correspondence)*, vol. CT-14, no. 1, pp. 103–104, Mar. 1967.

[195] KOMPIS, M. and N. DILLIER, "Simulating transfer functions in a reverberant room including source directivilty and head-shadow effects," *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2779 – 2787, May 1993.

[196] KOZEK, W., *Matched Weyl-Heisenberg expansions of non-stationary environments*, Ph. D. Thesis, Vienna Univ. Technol., Vienna, Austria, Mar. 1997, also appeared as Tech. Rep 96/I, NUHAG, Dept. Math., Univ. Vienna, Sept. 1996.

[197] KOZIN, F. and F. NAKAJIMA, "The order determination problem for linear time-varying AR models," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 2, pp. 250–257, Apr. 1980.

[198] KUNDUR, D. and D. HATZINAKOS, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, May 1996.

[199] KUNDUR, D. and D. HATZINAKOS, "Blind image deconvolution revisited," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 61–63, Nov. 1996.

[200] KUO, S. M. and J. KUNDURU, "Multiple reference subband adaptive noise canceler for hands-free cellular phone applications," *Journal of the Franklin Institute*, vol. 333B, no. 5, pp. 669–686, Sept. 1996.

[201] KUTTRUFF, H., *Room Acoustics*, Elsevier Applied Science Publishers, Essex, England, third edition, 1991.

[202] LAGENDIJK, R. L., J. BIEMOND, and D. E. BOEKEE, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Transactions on Speech, and Signal Processing*, vol. 38, no. 7, pp. 1180–1191, July 1990.

[203] LAGENDIJK, R. L., A. M. TEKALP, and J. BIEMOND, "Maximum likelihood image and blur identification: a unifying approach," *Optical Engineering*, vol. 29, no. 5, pp. 422–435, May 1990.

[204] LAMPARD, D. G., "Generalization of the Wiener-Khintchine theorem to nonstationary processes," *Journal Applied Physics*, vol. 25, no. 6, pp. 802–803, June 1954.

[205] LANG, S. W. and J. H. MCCLELLAN, "A simple proof of stability for all-pole linear prediction models," *Proc. IEEE*, vol. 67, no. 5, pp. 860–861, May 1979.

[206] LANGHANS, T. and H. W. STRUBE, "Speech enhancement by nonlinear multiband envelope filtering," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 156–159, Paris, France, May 1982.

[207] LANING, J. H. and R. H. BATTIN, *Random Processes in Automatic Control*, McGraw-Hill Book Co., Inc., New York, N.Y., 1956.

[208] LAVIELLE, M., "Optimal segmentation of random processes," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1365–1373, May 1998.

[209] LAW, N. F. and D. T. NGUYEN, "Multiple frame projection based blind deconvolution," *Electronic Letters*, vol. 31, no. 20, pp. 1734–1735, Sept. 1995.

[210] LAY, K. T. and A. K. KATSAGGELOS, "Image identification and restoration based on the expectation-maximization algorithm," *Optical Engineering*, vol. 29, no. 5, pp. 436–445, May 1990.

[211] LEE, C. K. and D. G. CHILDERS, "Cochannel speech separation," *Journal of the Acoustical Society of America*, vol. 83, no. 1, pp. 274–280, Jan. 1988.

[212] LEE, H. C., "Review of image-blur models in a photographic system using the principles of optics," *Optical Engineering*, vol. 29, no. 5, pp. 405–421, May 1990.

[213] LEVIN, M. J., "Instantaneous spectra and ambiguity functions," *IEEE Transactions on Information Theory*, vol. IT-10, pp. 95–97, Jan. 1964, correspondence.

[214] LI, H. T. and P. M. DJURIĆ, "MMSE parameter estimation of multiple chirp signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2606–2609, Atlanta, GA, USA, May 1996.

[215] LI, T. H., "Blind identification and deconvolution of linear systems driven by binary random sequences," *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 26–38, Jan. 1992.

[216] LI, T. H., "Estimation and blind deconvolution of autoregressive systems with nonstationary binary inputs," *Journal of Time Series Analysis*, vol. 14, no. 6, pp. 575–588, 1993.

[217] LI, T. H., "Blind deconvolution of linear-systems with multilevel nonstationary inputs," *The Annals of Statistics*, vol. 23, no. 2, pp. 690–704, 1995.

[218] LI, T. H. and K. MBAREK, "A blind equalizer for nonstationary discrete-valued signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 247–254, Jan. 1997.

[219] LI, W., J. C. H. POON, and W. C. SIU, "Recovery of single source signal from noisy and reverberant environment using second-order statistics," *Signal Processing*, vol. 75, no. 3, pp. 265–275, June 1999.

[220] LIGHTHILL, M. J., *Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press, 1958.

[221] LIM, J. S. and A. V. OPPENHEIM, "All-pole modeling of degraded speech," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-26, no. 3, pp. 197–210, June 1978.

[222] LIM, T. J., *System Identification Using Novel Adaptive Filter Structures*, Ph. D. Thesis, University of Cambridge, UK, Aug. 1995.

[223] LINDSAY, P. H. and D. A. NORMAN, *Human Information Processing; an introduction to psychology*, Academic Press, Inc., London, second edition, 1977.

[224] LINK, M. J. and K. M. BUCKLEY, "Robust real-time constrained hearing aid arrays," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 81–84, Mohonk Mountain House, New York, Oct. 1993.

[225] LIPORACE, L. A., "Linear estimation of nonstationary signals," *Journal of the Acoustical Society of America*, vol. 58, no. 6, pp. 1288–1295, Dec. 1975.

[226] LITOVSKY, R. Y., H. S. COLBURN, W. A. YOST, and S. J. GUZMAN, "The precedence effect," *Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[227] LIU, B. and P. A. FRANASZEK, "A class of time-varying digital filters," *IEEE Transactions on Circuit Theory*, vol. CT-16, no. 4, pp. 467–471, Nov. 1969.

[228] LIU, Q. G., B. CHAMPAGNE, and P. KABAL, "A microphone array processing technique for speech enhancement in a reverberant space," *Speech Communication*, vol. 18, no. 4, pp. 317–334, June 1996.

[229] LOYNES, R. M., "On the concept of the spectrum for non-stationary processes (with discussion)," *Journal of the Royal Statistical Society – Series B*, vol. 30, pp. 1 – 30, 1968.

[230] LUBMAN, D., "Fluctuations of sound with position in a reverberant room," *Journal of the Acoustical Society of America*, vol. 44, no. 6, pp. 1491–1502, Dec. 1968.

[231] LYON, R. H., "Progressive phase trends in multi-degree-of-freedom systems," *Journal of the Acoustical Society of America*, vol. 73, no. 4, pp. 1223–1228, Apr. 1983.

[232] LYON, R. H., "Range and frequency dependence of transfer function phase," *Journal of the Acoustical Society of America*, vol. 76, no. 5, pp. 1433–1437, Nov. 1984.

[233] MACKENZIE, J., J. HUOPANIEMI, V. VALIMAKI, and I. KALE, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39–41, Feb. 1997.

[234] MACLEAN, W. R., "On the acoustics of cocktail parties," *Journal of the Acoustical Society of America*, vol. 31, no. 1, pp. 79–80, Jan. 1959.

[235] MADDOX, J., "Cocktail party effect made tolerable," *Nature*, vol. 369, no. 6481, p. June, 517 1994.

[236] MAKHOUL, J., "Spectral analysis of speech by linear prediction," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-21, no. 3, pp. 140–148, June 1973.

[237] MAKHOUL, J., "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[238] MANN, H. B. and A. WALD, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, vol. 11, no. 3, 4, pp. 173 – 220, July-October 1943.

[239] MARGRAVE, G. F., "Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering," *Geophysics*, vol. 63, no. 1, pp. 244–259, Jan.-Feb. 1998.

[240] MARK, W. D., "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *Journal of Sound and Vibration*, vol. 11, no. 1, pp. 19–64, Jan. 1970.

[241] MARK, W. D., "Power spectral representation of nonstationary random processes defined over semi-infinite intervals," *Journal of the Acoustical Society of America*, vol. 59, no. 5, pp. 1184–1194, May 1976.

[242] MARMARELIS, V. Z., "A single-record estimator for correlation functions of nonstationary random processes," *Proc. IEEE*, vol. 69, no. 7, pp. 841–842, July 1981.

[243] MARMARELIS, V. Z., "Practical estimation of correlation functions of nonstationary Gaussian processes," *IEEE Transactions on Information Theory*, vol. IT-29, no. 6, pp. 937–938, Nov. 1983.

[244] MARQUES, J. S. and L. B. ALMEIDA, "Frequency-varying sinusoidal modeling of speech," *IEEE Transactions on Speech, and Signal Processing*, vol. 37, no. 5, pp. 763–765, May 1989.

[245] MARRO, C., Y. MAHIEUX, and K. U. SIMMER, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.

[246] MATZ, G., F. HLAWATSCH, and W. KOZEK, "Generalized evolutionary spectral analysis and the Weyl spectrum of non-stationary random processes," *IEEE Transactions on Signal Processing*, vol. 45, no. 6, pp. 1520–1534, June 1997.

[247] MAXWELL, J. A. and P. M. ZUREK, "Reducing acoustic feedback in hearing aids," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 304–313, July 1995.

[248] MCAULAY, R. J. and T. F. QUATIERI, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 744–754, Aug. 1986.

[249] MIDDLETON, D., "A statistical theory of reverberation and similar first-order scattered fields. Part I: Waveforms and the general process," *IEEE Transactions on Information Theory*, vol. IT-13, no. 3, pp. 372–392, July 1967.

[250] MIDDLETON, D., "A statistical theory of reverberation and similar first-order scattered fields. Part II: Moments, spectra, and special distributions," *IEEE Transactions on Information Theory*, vol. IT-13, no. 3, pp. 393–414, July 1967.

[251] MILLER, K. S. and L. A. ZADEH, "Solution of an integral equation occurring in the theories of prediction and detection," *IRE Transactions on Information Theory*, vol. IT-2, pp. 72–72, June 1956.

[252] MITCHELL, O. M. N., C. A. ROSS, and G. H. YATES, "Signal processing for a cocktail party effect," *Journal of the Acoustical Society of America*, vol. 50, no. 2, pp. 656–660, Aug. 1971.

[253] MIYOSHI, M. and Y. KANEDA, "Inverse filtering of room acoustics," *IEEE Transactions on Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[254] MOLGEDEY, L. and H. G. SCHUSTER, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, June 1994.

[255] MORGAN, D. P., E. B. GEORGE, L. T. LEE, and S. M. KAY, "Cochannel speaker separation by harmonic enhancement and suppression," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 407–424, Sept. 1997.

[256] MOURJOPOULOS, J. N., *The Removal of Reverberation from Signals*, Ph. D. Thesis, University of Southampton, UK, 1984.

[257] MOURJOPOULOS, J. N., "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217–228, Sept. 1985.

[258] MOURJOPOULOS, J. N., "Digital equalization methods for audio systems," *Journal of the Audio Engineering Society*, vol. 36, p. 384, May 1988, preprint 2598. Presented at the 84th Convention of the Audio Engineering Society, 1988 March 1-4, Paris.

[259] MOURJOPOULOS, J. N., "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994, preprint 3288. Presented at the 92nd Convention of the Audio Engineering Society, 1992 March 24-27, Vienna.

[260] MOURJOPOULOS, J. N., P. M. CLARKSON, and J. K. HAMMOND, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1858–1861, Paris, France, May 1982.

[261] MOURJOPOULOS, J. N. and J. K. HAMMOND, "Modelling and enhancement of reverberant speech using an envelope convolution method," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1144–1147, Boston, MA, USA, Apr. 1983.

[262] MOURJOPOULOS, J. N. and M. A. PARASKEVAS, "Pole and zero modeling of room transfer functions," *Journal of Sound and Vibration*, vol. 146, no. 2, pp. 281 – 302, Apr. 1991.

[263] MOURJOPOULOS, J. N., A. E. TSOPANGLOU, and N. D. FAKOTAKIS, "A vector quantization approach for room transfer function classification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3593–3596, Toronto, Ont., Canada, Apr. 1991.

[264] MRAD, R. B., S. D. FASSOIS, and J. A. LEVITT, "A polynomial-algebraic method for non-stationary TARMA signal analysis – Part I: The method," *Signal Processing*, vol. 65, no. 1, pp. 1–19, Feb. 1998.

[265] MUKHOPADHYAY, S. and P. SIRCAR, "Parametric modelling of non-stationary signals: A unified approach," *Signal Processing*, vol. 60, no. 2, pp. 135–152, July 1997.

[266] MULGREW, B., "Nonlinear filters and time series prediction," in *1st European Conference on Signal prediction and Analysis*, pp. 47–52, Prague, June 1997.

[267] MUNSHI, A. S., "Equalizability of room acoustics," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 217–220, San Francisco, CA, USA, Mar. 1992.

[268] MURON, O. and J. SIKORAV, "Modelling of reverberators and audioconference rooms," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 921–924, Tokyo, Japan, Apr. 1986.

[269] MURTHI, M. N. and B. D. RAO, "All-pole modeling of speech based on the minimum variance distortionless response spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 221–239, May 2000.

[270] NARENDRA, K. S., "Integral transforms for a class of time-varying linear systems," *IRE Transactions on Automatic Control*, vol. AC-6, pp. 311–319, Sept. 1961.

[271] NATHAN, K. S., Y. T. LEE, and H. F. SILVERMAN, "A time-varying analysis method for rapid transitions in speech," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 815–824, Apr. 1991.

[272] NAYLOR, A. W., "Generalized frequency response concepts for time-varying, discrete–time linear systems," *IEEE Transactions on Circuit Theory*, vol. CT-10, no. 5, pp. 428–440, Sept. 1963.

[273] NAYLOR, J. A. and S. F. BOLL, "Techniques for suppression of an interfering talker in co-channel speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 205–208, Dallas, TX, USA, Apr. 1987.

[274] NEELY, S. T. and J. B. ALLEN, "Invertibility of a room impulse response," *Journal of the Acoustical Society of America*, vol. 66, no. 1, pp. 165–169, July 1979.

[275] NOGA, J. L., *Bayesian State-Space Modelling of Spatio-Temporal Non-Gaussian Radar Returns*, Ph. D. Thesis, University of Cambridge, UK, Dec. 1998.

[276] OMURA, M., M. YADA, H. SARUWATARI, S. KAJITA, K. TAKEDA, and F. ITAKURA, "Compensating of room acoustic transfer functions affected by change of room temperature," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 941–944, Phoenix, Arizona, USA, Mar. 1999.

[277] ONO, M., K. ARAKAWA, M. MORI, T. SUGIMOTO, and H. HARASHIMA, "Separation of fine crackles from vesicular sounds by a nonlinear digital filter," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 2, pp. 286–291, Feb. 1989.

[278] OPPENHEIM, A. V., "Generalized superposition," *Information and Control*, vol. 11, no. 5-6, pp. 528–536, Nov.–Dec. 1967.

[279] OPPENHEIM, A. V., G. E. KOPEC, and J. M. TRIBOLET, "Signal analysis by homomorphic prediction," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-24, no. 4, pp. 327–332, Aug. 1976.

[280] OPPENHEIM, A. V. and J. S. LIM, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[281] OPPENHEIM, A. V. and R. W. SCHAFER, *Digital Signal Processing*, Prentice Hall, Inc., Englewood Cliff, NJ, 1975.

[282] OPPENHEIM, A. V., R. W. SCHAFER, and T. G. STOCKHAM, JR., "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, no. 8, pp. 1264–1291, Aug. 1968.

[283] O'RUANAIDH, J. J. K. and W. J. FITZGERALD, *Numerical Bayesian Methods Applied to Signal Processing*, Springer-Verlag, New York, 1996.

[284] PAPOULIS, A., *The Fourier Integral and its applications*, McGraw-Hill Book Co., Inc., 1962.

[285] PAPOULIS, A., "Predictable processes and Wold's decomposition: A review," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-33, no. 4, pp. 933–938, Aug. 1985.

[286] PAPOULIS, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Co., Inc., third edition, 1991.

[287] PARSONS, T. W., "Separation of speech from interfering speech by means of harmonic selection," *Journal of the Acoustical Society of America*, vol. 60, no. 4, pp. 911–918, Oct. 1976.

[288] PEI, S. C. and J. F. KIANG, "Simple approach for a class of linear time-varying digital filters with generalised delay elements," *IEEE Transactions on Circuits and Systems*, vol. CAS-33, no. 5, pp. 552–555, May 1986.

[289] PLOMP, R., "Acoustical aspects of cocktail parties," *Acustica*, vol. 38, no. 3, pp. 186–191, Sept. 1977.

[290] PLOMP, R. and A. J. DUQUESNOY, "Room acoustics for the aged," *Journal of the Acoustical Society of America*, vol. 68, no. 6, pp. 1616–1621, Dec. 1980.

[291] PORTER, D. and D. S. G. STIRLING, *Integral equations*, Cambridge University Press, 1990.

[292] PORTNOFF, M. R., "Implementation of the digital phase vocoder using the fast Fourier transform," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-24, no. 3, pp. 243–248, June 1976.

[293] PORTNOFF, M. R., "Time-frequency representation of digital signal and systems based on short-time Fourier analysis," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-28, no. 1, pp. 55–69, Feb. 1980.

[294] PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY, *Numerical Recipes in C*, Cambridge University Press, second edition, 1992.

[295] PRIESTLEY, M. B., "Evolutionary spectra and non-stationary processes (with discussion)," *Journal of the Royal Statistical Society – Series B*, vol. 27, pp. 204 – 237, 1965.

[296] PRIESTLEY, M. B., *Non-linear and Non-stationary Time Series Analysis*, Academic Press, Inc., 1988.

[297] PRIESTLEY, M. B. and H. TONG, "On the analysis of bivariate non-stationary processes," *Journal of the Royal Statistical Society – Series B*, vol. 35, pp. 135–166, 1973.

[298] PUNSKA, O., C. ANDRIEU, A. DOUCET, and W. J. FITZGERALD, "Bayesian segmentation of piecewise constant autoregressive processes using MCMC methods," Technical Report CUED/F-INFENG/TR. 344, Department of Engineering, University of Cambridge, UK, 1999.

[299] PUTNAM, W., D. ROCCHESSO, and J. SMITH, "A numerical investigation of the invertibility of room transfer functions," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 249–252, Mohonk Mountain House, New York, Oct. 1995.

[300] QUATIERI, T. F. and R. G. DANISEWICZ, "An approach to co-channel talker interference suppression using a sinusoidal model for speech," *IEEE Transactions on Speech, and Signal Processing*, vol. 38, no. 1, pp. 56–69, Jan. 1990.

[301] QUATIERI, T. F. and R. J. MCAULAY, "Audio signal processing based on sinusoidal analysis/synthesis," in Kahrs and Brandenburg [172], pp. 343–416.

[302] QUINN, A., "Regularised signal identification using Bayesian techniques," in A. Procházha, J. Uhliř, P. J. W. Rayner, and N. G. Kingsbury, editors, *Signal Analysis and Prediction*, pp. 151– 162, Birkhäuser, Boston, 1998.

[303] RABINER, L. R. and B. GOLD, *Theory and Application of Digital Signal Processing*, Prentice Hall, Inc., Englewood Cliff, NJ, 1975.

[304] RADLOVIĆ, B. D. and R. A. KENNEDY, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[305] RADLOVIĆ, B. D., R. C. WILLIAMSON, and R. A. KENNEDY, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

[306] RAJAN, J. J. and P. J. W. RAYNER, "Parameter estimation of time-varying autoregressive models using the Gibbs sampler," *Electronic Letters*, vol. 31, no. 13, pp. 1035–1036, June 1995.

[307] RAJAN, J. J. and P. J. W. RAYNER, "Generalised feature extraction for time-varying autoregressive models," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2498–2507, Oct. 1996.

[308] RAJAN, J. J., P. J. W. RAYNER, and S. J. GODSILL, "Bayesian approach to parametric estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 144, no. 4, pp. 249–256, Aug. 1997.

[309] RAO, A. and R. KUMARESAN, "On decomposing speech into modulated components," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 240–254, May 2000.

[310] RAO, S. and W. A. PEARLMAN, "Analysis of linear prediction, coding, and spectral estimation from subbands," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1160–1178, July 1996.

[311] RAY, G. C., "An algorithm to separate nonstationary part of a signal using midprediction filter," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2276–2279, Sept. 1994.

[312] RIHACZEK, A. W., "Signal energy distribution in time and frequency," *IEEE Transactions on Information Theory*, vol. IT-14, no. 3, pp. 369–374, May 1968.

[313] ROTSTEIN, H. and S. RAZ, "Gabor transform of time-varying systems: Exact representation and approximation," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, Apr. 1999.

[314] SALEH, B. E. A. and N. S. SUBOTIC, "Time-variant filtering of signals in the mixed time-frequency domain," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-33, no. 6, pp. 1479–1485, Dec. 1985.

[315] SANTHANAM, B. and P. MARAGOS, "Demodulation of discrete multicomponent AM-FM signals using periodic algebraic separation and energy demodulation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 2409–2412, Munich, Germany, Apr. 1997.

[316] SANTHANAM, B. and P. MARAGOS, "Harmonic analysis and restoration of separation methods for periodic signal mixtures: Algebraic separation versus comb filtering," *Signal Processing*, vol. 69, no. 1, pp. 81–91, Aug. 1998.

[317] SAUNDERS, G. H. and J. M. KATES, "Speech intelligibility enhancement using hearing-aid array processing," *Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1827–1837, Sept. 1997.

[318] SAYEED, A. M. and D. L. JONES, "Equivalence of generalized joint signal representations of arbitrary variables," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2959–2970, Dec. 1996.

[319] SCHELL, J. A., S. RITER, and R. K. CAVIN, "Dereverberation by linear systems techniques," *IEEE Transactions on Geoscience Electronics*, vol. Ge-9, no. 1, pp. 28–34, Jan. 1971.

[320] SEZAN, M. I. and A. M. TEKALP, "Survey of recent developments in digital image restoration," *Optical Engineering*, vol. 29, no. 5, pp. 393–404, May 1990.

[321] SHAH, S. I., L. F. CHAPARRO, and A. S. KAYHAN, "Evolutionary maximum entropy spectral analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. IV, pp. 285–288, Adelaide, SA, Australia, Apr. 1994.

[322] SHALVI, O. and E. WEINSTEIN, "System identification using nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 2055–2063, Aug. 1996.

[323] SHEPHARD, N., "Local scale models: State space alternative to integrated GARCH processes," *Journal of Econometrics*, vol. 60, pp. 181–202, 1994.

[324] SHINBROT, M., "A generalization of a method for the solution of the integral equation arising in optimization of time-varying linear systems with nonstationary inputs," *IRE Transactions on Information Theory*, vol. IT-3, pp. 220–224, Dec. 1957.

[325] SHINBROT, M., "On the integral equation occurring in optimization theory with nonstationary inputs," *Journal of Mathematics and Physics*, vol. 36, pp. 121–129, July 1957.

[326] SHON, S. and K. MEHROTRA, "Performance comparison of autoregressive estimation methods," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 14.3.1–14.3.4, San Diego, CA, USA, Mar. 1984.

[327] SILVA, F. M. and L. B. ALMEIDA, "Speech separation by means of stationary least-squares harmonic estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 809–812, Albuquerque, NM, USA, Apr. 1990.

[328] SILVERMAN, R. A., "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. IT-13, pp. 579–587, Sept. 1957.

[329] SIRCAR, P. and S. MUKHOPADHYAY, "Accumulated moment method for estimating parameters of the complex exponential signal models in noise," *Signal Processing*, vol. 45, no. 2, pp. 231–243, Aug. 1995.

[330] SIRCAR, P. and S. SHARMA, "Complex FM signal model for non-stationary signals," *Signal Processing*, vol. 57, no. 3, pp. 283–304, Mar. 1997.

[331] SIRCAR, P. and M. S. SYALI, "Complex AM signal model for non-stationary signals," *Signal Processing*, vol. 53, no. 1, pp. 35–35, Aug. 1996.

[332] SNYDER, R. L., "A partial spectrum approach to the analysis of quasi-stationary time-series," *IEEE Transactions on Information Theory*, vol. IT-3, no. 4, pp. 579–587, Oct. 1967.

[333] SPENCER, P. S., *System Identification with Application to the Restoration of Archived Gramophone Recordings*, Ph. D. Thesis, University of Cambridge, UK, June 1990.

[334] SPENCER, P. S. and P. J. W. RAYNER, "Separation of stationary and time-varying systems and its application to the restoration of gramophone recordings," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 292–295, Portland, OR, USA, May 1989.

[335] STARK, H. and J. W. WOODS, *Probability, Random Processes and Estimation Theory for Engineers*, Prentice Hall, Inc., Upper Saddle River, New Jersey, second edition, 1994.

[336] STEINBERG, B. D., "Note on "Correlation functions and power spectra in variable networks"," *Proc. IRE (Correspondence)*, vol. 40, p. 103, Jan. 1952.

[337] STOCKHAM, JR., T. G., T. M. CANNON, and R. B. INGRBRESTSEN, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, no. 4, pp. 678 – 692, Apr. 1975.

[338] STRICKLAND, R. N., "Transforming images into block stationary behavior," *Applied Optics*, vol. 22, no. 10, pp. 1462–1473, May 1983.

[339] STUBBERUD, A. R., *Analysis and Synthesis of Linear Tine-Variable Systems*, University of California Press, Berkeley and Los Angeles, 1964.

[340] STUBBS, R. J. and Q. SUMMERFIELD, "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 84, no. 4, pp. 1236–1249, Oct. 1988.

[341] STUBBS, R. J. and Q. SUMMERFIELD, "Algorithms for separating the speech of interfering talkers: Evaluations with voiced sentences and normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 87, no. 1, pp. 359–372, Jan. 1990.

[342] STUBBS, R. J. and Q. SUMMERFIELD, "Effects of signal-to-noise ratio, signal periodicity, and degree of hearing impairment on the performance of voice-separation algorithms," *Journal of the Acoustical Society of America*, vol. 89, no. 3, pp. 1383–1393, Mar. 1991.

[343] SUBBA RAO, T., "The fitting of nonstationary time-series models with time-dependent parameters," *Journal of the Royal Statistical Society – Series B*, vol. 32, no. 2, pp. 312–322, 1979.

[344] SUBRAMANIAM, S., A. P. PETROPULU, and C. WENDT, "Cepstrum-based deconvolution for speech dereverberation," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 5, pp. 392–396, Sept. 1996.

[345] SVEAN, J., A. KROKSTAD, and S. SØRSDAL, "An all digital concha hearing aid," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 85–88, Mohonk Mountain House, New York, Oct. 1993.

[346] TAMVACLIS, C., "Filtration of separable finite discrete signals," *IEE Proceedings, Part F: Communications, Radar and Signal Processing*, vol. 128, no. 1, pp. 1–8, Feb. 1981.

[347] TAMVACLIS, C., *Separability of Finite Discrete Signals*, Ph. D. Thesis, University of Cambridge, UK, 1981.

[348] TAN, S. L. and T. R. FISCHER, "Linear prediction of subband signals," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 9, pp. 1576–1583, Dec. 1994.

[349] TEKALP, A. M. and H. KAUFMAN, "On statistical identification of a class of linear space-invariant image blurs using non-minimum-phase ARMA models," *IEEE Transactions on Speech, and Signal Processing*, vol. 36, no. 8, pp. 1360–1363, Aug. 1988.

[350] TEKALP, A. M., H. KAUFMAN, and J. W. WOODS, "Fast recursive estimation of the parameters of a space-varying autoregressive image model," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-33, no. 2, pp. 469–472, Apr. 1985.

[351] TEKALP, A. M., H. KAUFMAN, and J. W. WOODS, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 963–972, Aug. 1986.

[352] THEOBALD, B., S. COX, G. CAWLEY, and B. MILNER, "Fast method of channel equalisation for speech signals and its implementation on a DSP," *Electronic Letters*, vol. 35, no. 16, pp. 1309–1311, Aug. 1999.

[353] THERRIEN, C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, Inc., 1992.

[354] THOMPSON, C., K. CHANDRA, and V. MEHTA, "Simulation of teleconferencing environments," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 131–136, Mohonk Mountain House, New York, Oct. 1995.

[355] TJØSTHEIM, D., "Spectral generating operators for non-stationary processes," *Advances in Applied Probability*, vol. 8, pp. 831–846, 1976.

[356] TOHYAMA, M. and R. H. LYON, "Transfer function phase and truncated impulse response," *Journal of the Acoustical Society of America*, vol. 86, no. 5, pp. 2025–2029, Nov. 1989.

[357] TOHYAMA, M. and R. H. LYON, "Zeros of a transfer function in a multi-degree-of-freedom vibrating system," *Journal of the Acoustical Society of America*, vol. 86, no. 5, pp. 1854–1863, Nov. 1989.

[358] TOHYAMA, M., R. H. LYON, and T. KOIKE, "Reverberant phase in a room and zeros in the complex frequency plane," *Journal of the Acoustical Society of America*, vol. 89, no. 4, pp. 1701–1707, Apr. 1991.

[359] TOHYAMA, M., R. H. LYON, and T. KOIKE, "Pulse waveform recovery in a reverberant condition," *Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2805–2812, May 1992.

[360] TOHYAMA, M., R. H. LYON, and T. KOIKE, "Source waveform recovery in a reverberant space by cepstrum dereverberation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 157–160, Minneapolis, USA, Apr. 1993.

[361] TOKUYAMA, H., "Evaluation of speech interference due to noise and reverberation," *Journal of Sound and Vibration*, vol. 127, no. 3, pp. 549–553, Dec. 1988.

[362] TONG, H., "On time-dependent linear transformations of non-stationary stochastic processes," *Journal of Applied Probability*, vol. 11, no. 1, pp. 53–62, Mar. 1974.

[363] TROUGHTON, P. T., *Simulation Methods for Linear and Nonlinear Time Series Models with Application to Distorted Audio Signals*, Ph. D. Thesis, University of Cambridge, UK, June 1999.

[364] TROUGHTON, P. T. and S. J. GODSILL, "Restoration of nonlinearly distorted audio using Markov Chain Monte Carlo methods," *Journal of the Audio Engineering Society*, vol. 46, no. 6, p. 569, June 1998, preprint 4679. Presented at the 104th Convention of the Audio Engineering Society.

[365] TSAO, Y. H., "Spectral model and time-varying covariance functions for the non-stationary processes," *Journal of the Acoustical Society of America*, vol. 76, no. 5, pp. 1422–1426, Nov. 1984.

[366] TSAO, Y. H., "Tests for nonstationarity," *Journal of the Acoustical Society of America*, vol. 75, no. 2, pp. 486–498, Feb. 1984.

[367] TSAO, Y. H., "Time-variant filtering for non-stationary random processes," *Journal of the Acoustical Society of America*, vol. 76, no. 4, pp. 1098–1113, Oct. 1984.

[368] TSAO, Y. H., "Uncertainty principle in frequency-time methods," *Journal of the Acoustical Society of America*, vol. 75, no. 5, pp. 1532–1540, May 1984.

[369] TSAO, Y. H. and J. K. HAMMOND, "Nonstationarity in acoustic fields," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 827–839, Sept. 1983.

[370] TSATSANIS, M. K. and G. B. GIANNAKIS, "Time-varying system identification and model validation using Wavelets," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3512–3523, Dec. 1993.

[371] TUFTS, D. W. and R. KUMARESAN, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 965 – 989, Sept. 1982.

[372] TURNER, C. H. M., "On the concept of an instantaneous power spectrum, and its relationship to the autocorrelation function," *Journal Applied Physics*, vol. 23, no. 11, pp. 1347–1351, Nov. 1954.

[373] VAIDYANATHAN, P. P., "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial." *Proc. IEEE*, vol. 78, no. 1, pp. 56–93, Jan. 1990, see *Proc. IEEE*, vol. 79, no. 2, pp. 242, Feb. 1991 for minor corrections.

[374] VAIDYANATHAN, P. P., "Multirate filtering and filter bands," in *IEEE Signal Processing Magazine* [19], pp. 17–19.

[375] VALAKAS, J., M. A. PARASKEVAS, and J. N. MOURJOPOULOS, "Pole-zero loudspeaker response analysis by bandsplitting and Prony modeling," in *Journal of the Audio Engineering Society (Abstracts)*, May 1996, preprint 4214. Presented at the 100th Convention of the Audio Engineering Society, Copenhagen, Denmark.

[376] VAN DE KERKHOF, L. M. and W. J. W. KITZEN, "Tracking of a time-varying acoustic impulse response by an adaptive filter," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1285–1294, June 1992.

[377] VASEGHI, S. V. and P. J. W. RAYNER, "The effects of non-stationary signal characteristics on the performance of adaptive audio restoration systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 377–380, Glasgow, UK, May 1989.

[378] VERMAAK, J., *Bayesian Modelling and Enhancement of Speech Signals*, Ph. D. Thesis, University of Cambridge, UK, Mar. 2000.

[379] VERMAAK, J., C. ANDRIEU, A. DOUCET, and S. J. GODSILL, "Non-stationary Bayesian modelling and enhancement of speech signals," Technical Report CUED/F-INFENG/TR. 351, Department of Engineering, University of Cambridge, UK, June 1999, submitted to IEEE Transactions on Speech and Audio Processing.

[380] VERMAAK, J., C. ANDRIEU, A. DOUCET, and S. J. GODSILL, "Bayesian model selection of autoregressive processes," Technical Report CUED/F-INFENG/TR. 360, Department of Engineering, University of Cambridge, UK, Jan. 2000.

[381] VERMAAK, J., M. NIRANJAN, and S. J. GODSILL, "An improved speech production model for voiced speech utilising a seasonal AR-AR model and Markov Chain Monte Carlo simulation," Technical Report CUED/F-INFENG/TR. 325, Department of Engineering, University of Cambridge, UK, June 1998.

[382] WADE, G., *Signal Coding and Processing*, Cambridge University Press, second edition, 1994.

[383] WANG, D. L. and J. S. LIM, "The unimportance of phase in speech enhancement," *IEEE Transactions on Speech, and Signal Processing*, vol. ASSP-30, no. 4, pp. 679–681, Aug. 1982.

[384] WANG, H. and F. ITAKURA, "An approach to dereverberation using multimicrophone sub-band envelope estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 953–956, Toronto, Ont., Canada, May 1991.

[385] WANG, H. and F. ITAKURA, "Dereverberation of speech signals based on sub-band envelope estimation," *IEICE Transactions on*, vol. E74, no. 11, pp. 3576–3583, Nov. 1991.

[386] WANG, H. and F. ITAKURA, "Realization of acoustic inverse filtering through multimicrophone sub-band processing," *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E75-A, no. 11, pp. 1474–1483, Nov. 1992.

[387] WATKINS, A. J. and N. J. HOLT, "Effects of a complex reflection on vowel identification," *Acustica*, vol. 86, pp. 532–542, 2000.

[388] WAX, M., "Order selection for AR models by predictive least squares," *IEEE Transactions on Speech, and Signal Processing*, vol. 36, no. 4, pp. 581–588, Apr. 1988.

[389] WEINSTEIN, E., M. FEDER, and A. V. OPPENHEIM, "Multi-channel signal separation by decorrelation," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 405–413, Oct. 1993.

[390] WEISS, S., G. W. RICE, and R. W. STEWART, "Multichannel equalization in subbands," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 203–206, Mohonk Mountain House, New York, Oct. 1999.

[391] WEXLER, J. and S. RAZ, "Discrete Gabor expansions," *Signal Processing*, vol. 21, no. 3, pp. 207–220, Nov. 1990.

[392] WHITE, L. B. and B. BOASHASH, "Cross spectral analysis of nonstationary processes," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 830–835, July 1990.

[393] WHITTLE, P., "Recursive relations for predictors of non-stationary processes," *Journal of the Royal Statistical Society – Series B*, vol. 27, no. 3, pp. 523–532, 1965.

[394] WIENER, N., *The interpolation, extrapolation and smoothing of stationary time series*, John Wiley, New York, 1949, *NDRC Report*, Cambridge, 1942.

[395] WIERWILLE, W. W., "A new approach to the spectrum analysis of nonstationary signals," *IEEE Transactions on Applications and Industry*, vol. 82, pp. 322–327, Nov. 1963.

[396] WOODRUFF, B. D. and D. A. PREVES, "Fixed filter implementation of feedback cancellation for in-the-ear hearing aids," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 22–23, Mohonk Mountain House, New York, Oct. 1995.

[397] WOODS, J. W., "Two-dimensional discrete Markovian fields," *IEEE Transactions on Information Theory*, vol. IT-18, no. 2, pp. 232–240, Mar. 1972.

[398] WOODS, J. W., "Markov image modeling," *IEEE Transactions on Automatic Control*, vol. AC-23, no. 5, pp. 846–850, Oct. 1978.

[399] WU, P. W. and H. LEV-ARI, "Optimized estimation of moments for nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1210–1221, May 1997.

[400] WU, W. R. and P. C. CHEN, "Subband Kalman filtering for speech enhancement," *IEEE Transactions on Circuits and Systems II – Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1072–1083, Aug. 1998.

[401] WYRSCH, S. and A. KAELIN, "A DSP implementation of a digital hearing aid with recruitment of loudness compensation and acoustic echo cancellation," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New York, Oct. 1997.

[402] XIA, X. G., "System identification using chirp signals and time-variant filters in the joint time-frequency domain," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2072–2084, Aug. 1997.

[403] YAMADA, H., H. WANG, and F. ITAKURA, "Recovering of broadband reverberant speech signal by sub-band MINT method," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 967–972, Toronto, Ont., Canada, May 1991.

[404] YEGNANARAYANA, B. and P. S. MURTHY, "Enhancement of reverberant speech using LP residual signal," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 267–281, May 2000.

[405] YOKOYAMA, Y., M. KUMAZAWA, Y. IMANISHI, and N. MIKAMI, "A new method of nonstationary time series analysis based on inhomogeneous AR equation," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2130–2136, Aug. 1997.

[406] YOKOYAMA, Y., M. KUMAZAWA, and N. MIKAMI, "Estimation of the AR order of an inhomogeneous AR model with input expanded by a set of basis," *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E83-A, no. 3, pp. 551–557, Mar. 2000.

[407] YOULA, D. C., "The solution of a homogeneous Wiener-Hopf integral equation occuring in the expansion of second-order stationary random processes." *IRE Transactions on Information Theory*, vol. IT-13, pp. 187–193, Sept. 1957.

[408] YOUNG, N., *An introduction to Hilbert Space*, Cambridge University Press, 1988.

[409] ZADEH, L. A., "Correlation functions and power spectra in variable networks," *Proc. IRE*, vol. 38, pp. 1342–1345, Nov. 1950.

[410] ZADEH, L. A., "The determination of the impulsive response of variable networks," *Journal Applied Physics*, vol. 21, pp. 642–645, July 1950.

[411] ZADEH, L. A., "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, pp. 291–299, Mar. 1950.

[412] ZADEH, L. A., "Correlation functions and spectra of phase and delay-modulated signals," *Proc. IRE*, vol. 39, pp. 425–428, Apr. 1951.

[413] ZADEH, L. A., "General input-output relations for linear networks," *Proc. IRE (Correspondence)*, vol. 40, p. 103, Jan. 1952.

[414] ZADEH, L. A., "A general theory of linear signal transmission systems," *Journal of the Franklin Institute*, vol. 253, pp. 293–312, 1952.

[415] ZADEH, L. A., "On the theory of filtration of signals," *Zeitschrift fur Angenwante Mathematix und Physick*, vol. 111, pp. 149–156, 1952.

[416] ZADEH, L. A., "On a class of stochastic operators," *Journal of Mathematics and Physics*, vol. 32, pp. 48–53, Apr. 1953.

[417] ZADEH, L. A., "Time-varying networks, I," *Proc. IRE*, vol. 49, pp. 1488–1503, Oct. 1961.

[418] ZADEH, L. A. and K. S. MILLER, "Generalised ideal filters," *Journal Applied Physics*, vol. 23, pp. 223–228, 1952.

[419] ZADEH, L. A. and J. R. RAGAZZINI, "An extension of Wiener's theory of prediction," *Journal Applied Physics*, vol. 21, pp. 645–655, July 1950.

[420] ZOU, M. Y. and R. UNBEHAUEN, "An algebraic theory for separation of periodic signals," *Archiv fur Elektronik und Ubertragungstechnik – International Journal of Electronics and Communications*, vol. 45, no. 6, pp. 351–358, Nov. – Dec. 1991.

Author Index

Page numbers in **bold text** indicate the page of the bibliography entry.

Index